

Frictions in the Social and Economic Lives of Underprivileged People

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Natalia Ordaz Reynoso

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Paul Glewwe, Advisor
Marc Bellemare, Co-Advisor

May, 2021

© Natalia Ordaz Reynoso 2021
ALL RIGHTS RESERVED

Acknowledgements

I am grateful for the support and guidance of the members of my Committee: Professor Paul Glewwe, Professor Marc Bellemare, Professor Jason Kerwin and Professor Audrey Dorelien. This dissertation would not have been the same without their careful and insightful feedback at each stage of the process. I would like to emphasize the acknowledgement to Professor Kerwin, as the third chapter of this dissertation stems from co-authored work.

I also greatly benefited throughout the course of my PhD from discussions with peers in my program, in the Kerwin-Dorelien PhD Lab, the Bellemare Research Group, and the Economics department.

I would like to thank Professor Joseph Ritter, whose class I enjoyed very much, and who gave me the first comments on what would become my job market paper. I am also grateful to all my graduate school instructors, who enthusiastically showed me what they find interesting and exciting about Economics.

Finally, I would like to acknowledge that without the support of my network, especially, my parents, I would not have had the opportunity to dedicate six years of my life to learn about what I find interesting. This is a privilege I do not take for granted.

Dedication

To my family.

To my mom, who made sure I had the tools I would need to have better opportunities, and who also emphasized how important it is to work so that others can have them too. Ma, if you hadn't enrolled me in English language school, none of this would be possible. Thank you for always thinking about what was best for me, and thank you for teaching me to care about others.

To my dad, who was the first coach of my critical thinking. Pa, thank you for teaching me about the importance of thinking and learning. Thank you for teaching me that ideas are important, and for encouraging me to decide which ones I believe in. You are my ultimate mentor.

To my husband, Pedro, who supports me in all my endeavors. I love you.

Abstract

In the last century, standards of living around the world have improved. However, this progress has not been equal across, nor within countries. This dissertation consists of three chapters that aim to contribute to answering the question of why this has been the case, in three separate contexts.

Chapter 1: I test whether the implementation of the California Paid Family Leave Act increased young women's human capital investment, specifically college enrollment. Using a synthetic control approach, I estimate that the policy increased the probability that women enroll in college by about 2 percentage points. This effect is statistically significant at the 5% level and persists for at least several years. I present a simple human capital model of women's schooling choices that characterizes these results as the effect of an expected decrease in the effects of motherhood on labor supply. Finally, I present evidence from survey data and Internet searches that provides support to the hypothesized mechanism: women are more likely to enroll in college because they expect that the policy will increase their future labor supply.

Chapter 2: Directly eliciting individuals' subjective beliefs via surveys is increasingly popular in social science research, but doing so via face-to-face surveys has an important downside: the interviewer's knowledge of the topic may spill over onto the respondent's recorded beliefs. Using a randomized experiment that used interviewers to implement an information treatment, we show that reported beliefs are significantly shifted by interviewer knowledge. Trained interviewers primed respondents to use the exact numbers used in the training, nudging them away from higher answers; recorded responses decreased by about 0.3 standard deviations of the initial belief distribution. Furthermore, respondents with stronger prior beliefs were less affected by interviewer knowledge.

Chapter 3: Governments across the world subsidize soup kitchen programs, but there is little evidence on whether these improve food security. I study a soup kitchen program funded by the Mexican government in 2013 to examine whether it has caused an improvement in food security. I find no mean municipal effects for six different measures of food security. I analyze a sub sample of the most food insecure and identify some positive effects within that sector of the population. My results suggest that the effect of the program on food security is concentrated in the lower end of the food security distribution, but challenge the assumption that subsidizing prepared food will mechanically improve mean food security significantly. I also estimate that the presence of soup kitchens in the most ex-ante food insecure municipalities decreases average food expenditures. This result points to diverse effects of soup kitchen programs.

Contents

Acknowledgements	i
Dedication	ii
Abstract	iii
List of Tables	vi
List of Figures	viii
1 Introduction	1
2 Paid Maternity Leave and Women’s Human Capital: Evidence from California	4
2.1 Introduction	4
2.2 Background	7
2.2.1 Maternity Leave	7
2.2.2 The California Paid Family Leave Act	8
2.3 Data and Empirical Strategy	9
2.3.1 Data and Descriptive Statistics	9
2.3.2 Empirical Strategy	10
2.4 Results	13
2.4.1 Main Results	13
2.4.2 Inference	15
2.4.3 Heterogeneity by Race and Ethnicity	16
2.5 Model	20
2.6 Other Evidence	24
2.6.1 Survey Data: General Social Survey	24
2.6.2 Google Trends	27
2.7 Conclusions	29

3	You Know What I Know: Interviewer Knowledge Effects in Subjective Expectation Elicitation	32
3.1	Introduction	32
3.2	Background	35
3.2.1	Subjective Expectations and Demographic Decisions	35
3.2.2	The HIV Epidemic in Malawi	36
3.2.3	Subjective Expectations About HIV in Malawi	37
3.3	Data and Empirical Design	38
3.4	Empirical Strategy	40
3.5	Results	41
3.5.1	Main Estimates	41
3.5.2	Alternative Explanations	44
3.6	Mechanisms	49
3.6.1	Priming	50
3.6.2	Encouraging Guesses	51
3.6.3	Interviewer Knowledge and Respondent Priors	53
3.7	Preventing and Mitigating Interviewer Knowledge Effects	55
3.8	Conclusion	57
4	Soup Kitchens and Food Security: The Case of Mexico’s Crusade Against Hunger	59
4.1	Introduction	59
4.2	Background	61
4.3	Data and Empirical Strategy	63
4.4	Results	68
4.4.1	Effects on Food Expenditure	74
4.4.2	Political Alignment	77
4.5	Conclusions	78
5	Conclusion	81
	References	82
	Appendix A. Appendix to Chapter 3	97

List of Tables

2.1	Synthetic Control Weights: College Enrollment	14
2.2	V Weights	14
2.3	Synthetic Control Balance	15
2.4	Value Functions	22
2.5	Triple Differences Estimation on Variables of Women’s Roles	27
3.1	Effects of interviewer knowledge on reported risk beliefs	43
3.2	Regression discontinuity estimates	46
3.3	Treatment Effects on Answering Exactly 10%	51
3.4	Priming by interviewers when initial answer was 50%	52
3.5	Heterogeneity in treatment effects	54
4.1	Measures of Food Security	64
4.2	Summary Statistics for Treated and Non-treated Households	66
4.3	Social Capital Proxy Questionnaire	68
4.4	Estimation of Equation 1 with Binary Treatment on Measures of Food Security	69
4.5	Estimation of Equation 1 with Binary Treatment on Measures of Food Security for the Restricted Sample	70
4.6	Estimation of Equation 1 with Continuous Treatment on Measures of Food Security	71
4.7	Estimation of Equation 1 with Continuous Treatment on Measures of Food Security for the Restricted Sample	72
4.8	Estimation of Equation 1 with Continuous Treatment on Quarterly Per Capita Expenditure on Food, in Mexican Pesos	76
4.9	Estimation of Equation 3 with Binary Treatment on Measures of Food Security for the Restricted Sample	79
A.1	HIV Transmission Risk Survey Questions (Female Versions)	98
A.2	Balance of Sexual Activity and Demographics	99
A.3	Effects of Interviewer Knowledge on Reported Risk Beliefs, Controlling for Interviewer and Sampling Strata Fixed Effects Only	100

A.4	Effects of Interviewer Knowledge on Reported Risk Beliefs, Controlling for Sampling Strata Fixed Effects Only	100
A.5	Distribution of Religious Denomination and Ethnic Group By Treatment Arm	101
A.6	Literacy Rates in Chichewa by Ethnic Group and Study Arm	102
A.7	Effects of Interviewer Knowledge on Reported Risk Beliefs, Adding Controls for Religion, Ethnicity, and Literacy	102
A.8	Treatment Effect Spillovers	103
A.9	Collinearity Diagnostics for Age, Years of Schooling, and Years Sexually Active	104
A.10	Heterogeneity in Treatment Effects by Different Measures of Schooling . . .	105
A.11	Correlations between Schooling Measures and HIV Risk Beliefs in the Control Group	106

List of Figures

2.1	College Enrollment Trends in California and Other States	11
2.2	College Enrollment Rates in California and Other States for the 1987 Cohort	12
2.3	College Enrollment Rates for California and Synthetic California	16
2.4	Histogram of RMSPE Ratios	17
2.5	High School Graduation Rates for California and Synthetic California . . .	18
2.6	Synthetic Control: College Enrollment of Asian Women	18
2.7	Synthetic Control: College Enrollment of Black Women	19
2.8	Synthetic Control: College Enrollment of Hispanic Women	19
2.9	Synthetic Control: College Enrollment of White Women	20
2.10	Case A	23
2.11	Case B	24
2.12	Case C	24
2.13	Case A With Distribution of Motherhood Costs	25
2.14	Filtered Google Search Index: “Working Mom”	29
2.15	Filtered Google Search Index: “Paid Maternity Leave California”	30
2.16	Filtered Google Search Index: “Paid Maternity Leave ”	30
3.1	Measured Risk Beliefs over Time, by Study Arm	41
3.2	Regression discontinuity plots	45
4.1	Distribution the Sum of all Food Security Questions at Baseline for Treated and Non-treated Households	67
4.2	Event Analysis for the Full Sample	73
4.3	Event Analysis for the Restricted Sample	74
4.4	Trends for the Sum of Food Security Dummy Variables	75
A.1	Timeline of Experiment	98

Chapter 1

Introduction

The world has seen a striking improvement in the standard of living in the last century. Nevertheless, progress has not been equal across, or even within, countries. Which socioeconomic frictions have prevented some groups from benefiting as much as others? How can we remove them?

There may not be one simple answer to the above question. I therefore focus on different margins separately. The ultimate goal is to understand what the frictions are to general progress so that we can remove them. In this dissertation, I study two policies designed to reduce such frictions: soup kitchens and paid maternity leave. In addition, I investigate a technical issue that often arises in the study of economic development: problems eliciting subjective beliefs, which influence decision-making under uncertainty.

The first margin that I focus on is gender and the labor market. The gap between women's and men's wages and employment has been widely documented. While the gap has been narrowed on several dimensions, it is persistent. A smaller share of women than men participate in the labor force, and women earn less on average. Economists and other social scientists have investigated the persistence of this gap, and some policies have been put in place or redesigned to try to reduce it. One such policy is paid maternity leave.

Social scientists have found evidence of different effects of paid maternity leave programs. When they receive at least partial pay, women take longer maternity leave than when it is unpaid (Baum and Ruhm, 2016), and women's employment, hours worked and wages after giving birth increase as a result of these policies (Dustmann and Schönberg (2012), Rossin-Slater et al. (2013)).

In Chapter 2, I study the effect of the California Paid Family Leave Act on women's schooling. If women anticipate that paid maternity leave increases job retention after having a child, the expected benefit of schooling increases, as women would earn benefits from their human capital over a longer period. Another possibility is that women interpret the

implementation of these policies as a signal of more parent-friendly workplaces, decreasing the expected cost of motherhood, which in turn increases the likelihood that women continue working after having a child. I focus on the effect of paid maternity leave on college enrollment rates, using the synthetic control method of Abadie et al. (2010).

My results show that, for women, the policy is associated with a 1.8 percent-point increase in the probability of starting college. Studying indirect effects of maternity leave policies is key to understanding the determinants of labor market disparities, and the effects that policies may have on them. Furthermore, my results provide evidence that women's labor-force participation is hindered by the tradeoffs they face between motherhood and their careers, which is an important insight for policymakers.

In Chapter 3, I explore interviewer knowledge effects in subjective belief elicitation. Uncertainty is an intrinsic part of people's lives. When making decisions, people often do not know the objective probabilities of an outcome, but they have a set of beliefs about the likelihood of different scenarios. These beliefs influence the decisions they make. For this reason, researchers often want to study people's beliefs about the probabilities of specific events, particularly in health-related issues, where people, and even more so the poor, face substantial uncertainty.

Researchers have designed several ways to elicit these beliefs. This is usually done through in-person interviews. This chapter studies a potential problem with this method: the possibility that what the interviewer knows or believes may contaminate the respondents' answers. This work is coauthored with Professor Jason Kerwin, who performed the original experiment.

Interviewer effects have been documented in many contexts: participants give different answers to questions depending on characteristics of who is asking. Examples include interviewer's race and ethnicity (Adida et al. (2011), Cotter et al. (1982), and Dionne (2014)), gender (Becker et al. (1995), and McCombie and Anarfi (2002)) and social or cultural proximity (Weinreb (2006)).

In this chapter we study whether information held by interviewers contaminates participants' answers about their own beliefs. Interviewers that had received information about the true transmission risk of HIV elicited lower average baseline responses, closer to the true transmission risk. Informed interviewers appear to have primed respondents to use the exact numbers used in the information training, and nudged them away from higher answers. We find that recorded responses decrease by about 0.3 standard deviations of the initial belief distribution.

After identifying this effect, we suggest corrections from the perspectives of interviewer recruitment, survey design, experiment setup, and data analysis. This aids researchers in designing better instruments to study people's subjective beliefs, minimizing the risk

of contamination. Understanding biases in recorded risk beliefs is very important, since subjective beliefs are examined extensively in the study of economic development. If our measures of risk beliefs are biased, we will have a harder time studying the behaviors that depend on them, and some of these behaviors are very important to the study of economic development.

Preventing hunger and malnutrition has been a major concern for developed and developing countries. Increasingly so as the literature has found various negative effects of food insecurity. Malnutrition and food insecurity have short and long term effects on variables ranging from low productivity to decreased child health (Gundersen and Kreider, 2009) and reduced non-cognitive development (Howard, 2011).

Governments and international organizations expend large amounts of resources on programs aimed at improving food security, including soup kitchens. Programs operate under the assumption that setting up soup kitchens will automatically reduce overall food insecurity, yet there is little evidence to show whether these interventions have the intended effect. There may be barriers to food security aside from availability. For example, people may have work schedules that prevent them from attending at specific times, people may commute away from their place of residence or they may be simply unaware of the program.

In Chapter 4 of this dissertation I study whether a soup kitchen program, funded by the Mexican government, has improved local food security. I take a difference-in-differences approach using six different measures of food security, restricting the analysis to disadvantaged municipalities. I first find that the program is not associated with higher average municipal food security for disadvantaged municipalities. When analyzing only the most food insecure municipalities, I find positive effects on food security. These results suggest that the effect of the program is concentrated among the least food secure households.

My results challenge the assumption that subsidizing prepared food will automatically improve average food security. Still, my results are encouraging, as they show positive effects on the most food insecure, and a reduction in food expenditure for this group as well. This chapter stresses the importance of rethinking policy, as seemingly intuitive solutions may be missing key components of the underlying problem, and also highlights the fact that policymakers sometimes have different priorities than program beneficiaries themselves.

These three chapters share a common goal: better understanding barriers that underprivileged groups systematically face. Understanding these frictions is the main goal of my dissertation as a whole, and my research agenda for the future. Whether studying specific policies or contributing to improve research methods, my goal is to continue doing research that can help improve the lives of underprivileged groups. I want to continue working to answer questions that can inform policy to this end, either in an academic setting or within an international research organization.

Chapter 2

Paid Maternity Leave and Women's Human Capital: Evidence from California

2.1 Introduction

Women face important tradeoffs between work and family. Despite the increase in female labor supply in the last century, women still encounter several barriers to participation in the laborforce. These include lower salaries, barriers to promotion, and pregnancy related job loss. Maternity leave policies were originally established to provide some protection from gender discrimination in the workplace, so that women could have some time to bond with their children and recover from childbirth without risking their jobs. These policies have been promoted as a way to ease the tradeoffs between work and family that women face, especially given the cultural expectation that they take the role of primary caregivers for their children.

Economists and other social scientists have studied some implications of maternity leave policies. Particular attention has been paid to effects on women's labor market outcomes, including those on female labor supply, female labor earnings, and the gender gap in employment outcomes. Despite considerable interest in the effects of these policies on women's labor-market outcomes, effects on women's human capital investment decisions have not yet been studied. If these policies have impacts on women's labor-market outcomes, or if women believe they will, investments decisions should respond.

I study the question of whether paid maternity leave changes women's human capital investment decisions. I propose that women make investment decisions based on their expectations of the returns to those investments. Do women expect maternity leave to ease

the work-family tradeoffs that they face? Are women’s educational choices constrained by these tradeoffs?

I use the introduction of the California Paid Family Leave Act (CPFLA), which was the first of its kind in the US, to estimate the relationship between paid maternity leave and women’s human capital investment decisions. I use the synthetic control approach of Abadie and Gardeazabal (2003) to compare women in California who were likely to make their college enrollment decisions before and after the policy went into effect to two equivalent groups of women in the synthetic control. I estimate that the policy increased female college enrollment by about two percentage points. This effect is statistically significant and persists for at least several years.

Beyond the rationale that it reduces discrimination in the workplace, promoters of maternity leave argue that these policies are welfare increasing for mothers and children. There is evidence of improvements in child health outcomes, including increases in birth weight and a reduction in infant mortality among children of highly educated, married women (Rossin (2011)), as well as long-term benefits for children, such as higher wages and lower high-school dropout rates (Carneiro et al. (2015b)).

One literature has focused on the effects of maternity leave policies on women’s labor market outcomes and the gender gap. Researchers have found that mothers tend to extend their leave when the leave is paid (Baum and Ruhm (2013)), and that paid leave increases hours worked and wages (Rossin-Slater et al. (2011)). Women are more likely to be employed 9 to 12 months after giving birth (Baum and Ruhm (2013)), and overall job retention after having a child increases (Rossin-Slater et al. (2011)). However, Bailey et al. (2019) provide evidence that, over the long term, the CPFLA has not decreased the gender gap in labor force participation, and may have widened it for some groups. They also find lower employment levels for new mothers who used paid leave. The lack of apparent consensus in the literature regarding the effects of maternity leave policies on women’s labor-market outcomes stems in part from the fact that different features of maternity leave policies may have different impacts. For example, the selection effects of the job-protection and wage replacement features of paid leave policies may be in different directions (Rossin-Slater (2017), Stearns (2015)).

Despite this extensive interest in the effects of maternity leave policies on women’s labor supply, little attention has been paid to the effects of these policies on women’s human capital investment decisions. Expected effects of maternity leave policies on labor market outcomes may provide incentives for women to increase investments in their human capital. Women may interpret mandated paid maternity leave policies as a signal of more parent-friendly work environments or, more generally, as a decrease in the cost of motherhood on their labor supply. Women can then be expected to increase their labor supply, either

along the intensive or the extensive margins, relative to the counterfactual scenario of no such policy. Under these conditions, each additional unit of human capital becomes more profitable, as it would yield benefits over a longer period of time. In other words, if women expect to work more as a result of this policy, investing in their human capital is more profitable.

These effects can arise regardless of whether women actually remain in the labor-force after motherhood. The expectation that they are more likely to remain in the labor-force after having a child is enough to drive an increase in investment, in this case college enrollment. In other words, beyond what happens to labor supply ex-post, women make investment decisions using their ex-ante expectations of what will happen to their labor supply. Because at the time of implementation of the CPFLA, paid maternity leave policies were new in the US, there had not been much opportunity for women to update their beliefs by observing actual effects of paid leave on labor supply. In this sense, my results are not susceptible to the apparent lack of consensus in the literature about the labor-market effects of maternity leave, because the mechanism does not depend on the realized effects of the policy, but the ex-ante expectation that women have.

The contribution of this chapter is twofold. First, it shows that maternity leave policies can influence women’s schooling decisions, and thus shows that women’s schooling decisions are constrained by the work-family tradeoffs that they face. Second, this chapter brings attention to the fact that women expect these policies to ease such tradeoffs.

I present a human capital accumulation model, modified from Kuziemko et al. (2018), that rationalizes the mechanisms depicted above. The model describes women’s schooling decisions in the face of motherhood effects on their labor supply. Working with this model, I formalize my empirical results as women interpreting this policy as a sign of lower expected cost of motherhood on future labor supply. To further motivate my hypothesized mechanism, I present evidence from survey data that supports the theoretical model. I also present Google Trends data showing that interest in paid maternity leave increased in California (more than elsewhere) around the time of policy implementation, which provides further support to the hypothesized mechanism.

The rest of the chapter is structured as follows: Section 2 reviews the literature on the effects of maternity leave policies, and briefly describes the CPFLA. Section 3 describes the data and my empirical strategy. Section 4 provides the synthetic control estimates. Section 5 presents the model, Section 6 considers additional evidence that supports the hypothesized mechanism. Section 7 concludes.

2.2 Background

2.2.1 Maternity Leave

Maternity leave policies have the purpose of providing protection against workplace discrimination, and allow time for mothers to recover from childbirth, care for, and bond with a new child. These policies have received much interest from the research community and policy makers. Effects of maternity leave on a variety of outcomes have been studied. For example, there is substantial evidence that takeup increases when leave is paid or expanded (Han et al. (2009), Baum and Ruhm (2016), Bartel et al. (2018)).

There is also evidence of positive effects on outcomes for children. Increases in leave entitlement leads to a decline in high school dropout rates for children (Carneiro et al. (2015a)). Paid leave increases the likelihood and duration of breastfeeding (Pac et al. (2019)), and improves birth and health outcomes (Rossin (2011), Stearns (2015)). Bailey et al. (2019) find evidence that the CPFLA increased the time that mothers spend with their children.

Given labor market disparities between men and women, and gender disparities in time allocated to child rearing, there is an extensive literature that has focused on the labor market effects of maternity leave policies. Theoretical effects of maternity leave on post-leave employment outcomes are ambiguous (Klerman and Leibowitz (1998)), and there is an apparent lack of consensus in the empirical evidence.

On one hand, several studies have found that maternity leave is associated with an increase in employment after birth (Baum and Ruhm (2016), Kluve and Tamm (2013)), job attachment (Baum and Ruhm (2016), Byker (2016)), and hours and weeks worked (Rossin-Slater et al. (2013)). It is important to note that these effects are measured at different times. For example, effects may differ three years after birth vs. in the period directly before and after childbirth. The effects of these policies are plausibly different in different moments of the child’s life, and for different parities.

In contrast, Bailey et al. (2019) use a large administrative data set to evaluate the long term effects of the CPFLA. They find no evidence that the policy has closed the gender gap in labor-market outcomes such as employment, wage earnings or job attachment. In fact, they find a reduction in annual wages for new mothers who take leave via the CPFLA.

The apparent lack of consensus on the effects of maternity leave policies on women’s future labor market outcomes likely partially stems from the variation that exists in maternity leave policies and how these interact with the context (see Rossin-Slater (2017) for a comprehensive review of the literature on maternity leave policies). Different features of leave policies may have different effects. For example, FMLA offers job protection but no wage replacement, while CPFLA offers partial wage replacement but no job protection. These

two features may impact have different effects across different groups of women (Stearns (2015)). There is evidence of large heterogeneity in the effects across income groups (Bailey et al. (2019), Bedard and Rossin-Slater (2016)), and the effects may be very different depending on the length of the leave (Lequien (2012)).

The lack of conclusive evidence notwithstanding, to study the effects of women’s schooling decisions, what matters is what women expect to happen, ex-ante. In particular, what they expect at the time when they make their investment decisions. In the case of the CPFLA, women had little evidence on what would happen to their labor supply post motherhood, since this was the first mandated paid maternity leave policy implemented in the US.

In Kuziemko et al. (2018), it is shown that in the US and the UK, women have increased their schooling in the last few decades despite the fact that women are not working outside the home much more than before. The authors frame these two facts as women systematically underestimating the cost of motherhood on their labor supply. Women make schooling decisions expecting they will work after motherhood, but when they become mothers they receive an information shock and realize it is too costly to do both, so they end up with inefficiently high levels of human capital for the labor they end up supplying.

Kuziemko et al. (2018) argue that the reason for this systematic underestimation is that the cost of motherhood on labor supply has increased across cohorts, but women make decisions as if it had remained constant. The authors suggest that this increase is not driven by the work side of the work-family tradeoff, but by the family side, perhaps via increased requirements on childbearing (such as the promotion of breast feeding) or decreased support from networks. In this chapter I argue that, additionally, women may have taken the implementation of maternity leave policies as a sign of a decrease in the work side of the tradeoff, perhaps inferring from the implementation of maternity leave policies that workplaces were becoming more family friendly, or that society was more accomodating to women’s dual role as childrearsers and workers.

Regardless of the evidence we currently hold about the labor market effects on women’s labor supply, if, at the time they were making their schooling decisions, women expected that the CPFLA would increase their probability of returning to the work place after having a child, that would be enough to drive the effect on college enrollment I find in this chapter.

2.2.2 The California Paid Family Leave Act

Since 1993 and before the CPFLA passed, in 2004, some working women in the US were entitled to 12 weeks of job-protected leave per child (born to them or that they adopted) via the Family and Medical Leave Act (FMLA). This policy also provides job-protected

leave for other reasons, such as caring for a sick family member, and it also grants paternity leave to eligible fathers. Parents are eligible for 12 weeks of job protection through FMLA if they had worked for an eligible employer at least 1,250 hours in the 12 months before taking the leave. Eligible employers include those with over 50 employees that lived within 75 miles. Fathers and mothers are entitled to the same benefits via FMLA.

The California Paid Family Leave Act introduced partial pay benefits for eligible employees living and working in the state of California, that had a new child entering the family, or were taking care of a sick family member. Eligibility requirements include that workers have earned \$300 in wages that were subject to the State Disability Insurance deductions. The CPFLA does not provide job protection.

Those who are eligible for both FMLA and CPFLA can take both simultaneously, thus enjoying job protection via FMLA for up to twelve weeks, and (partial) wage replacement via CPFLA for up to 6 weeks. Even though CPFLA has fewer eligibility restrictions than FMLA, those women who are not eligible for FMLA may find it difficult to claim wage replacements via CPFLA if they are not covered by job-protection policies. It is also important to note that some women, especially those earning higher wages, may have already enjoyed paid maternity leave provided by their employers as part of a private benefit package.

2.3 Data and Empirical Strategy

2.3.1 Data and Descriptive Statistics

I use pooled Current Population Survey (CPS) data from all monthly CPS surveys from 2011 to 2017, through IPUMS-CPS.¹ The monthly CPS is a rotating panel, so from the pooled data I take the latest observation available for each woman. I drop from the sample women who were younger than 22 when they were last interviewed, to keep only women who have plausibly already finalized their decision of enrolling in college or not.

For privacy purposes, the CPS does not include information on participant's year of birth. To be able to perform aggregate analyses at the birth year-state level, I impute birth year by subtracting age at interview and year of interview. This can generate some noise in my imputed birth year variable, depending on month of birth and month of interview. However, this noise is not a major concern, since my imputation would not misclassify women by more than one calendar year, and the direction in which the misclassification occurs is not likely related to any relevant characteristics. As a robustness check, I perform

¹IPUMS-CPS, Sarah Flood, Miriam King, Steven Ruggles, and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 5.0. [dataset]. Minneapolis: University of Minnesota, 2017. <https://doi.org/10.18128/D030.V5.0>.

the main analyses defining treatment for birth-year cohorts one and two years before and after treatment actually occurred, and results are qualitatively similar.

As my outcome variable of interest, I define the binary variable “college enrolled”, which takes a value of one if the woman has a college degree or any college studies, and zero otherwise. This variable is constructed from different categories that women can select from in the CPS when reporting their educational attainment. Women who report having 1 year or college, some college but no degree, or anything beyond that count as college enrolled under this categorization. Other margins of human capital accumulation can be explored in a similar way, for example by using a binary variable for having attended graduate school as a dependent variable.

Other variables I use from the CPS include four race and ethnicity categories (Asian, Black, White, and Hispanic), high school graduation, and total household income. Using the described sample from the pooled CPS data, I collapse by state and the imputed birth year to obtain a strongly balanced panel of state and birth year averages of the variables I use.

I abstract from fertility choices, by including women in my sample regardless of whether they have children or not. I do this in part because there is reason to believe that the timing of pregnancy or motherhood can be susceptible to the policy, and in part because selection into motherhood itself cannot be properly dealt with using this method. Additionally, if I wanted to ensure that women had plausibly completed their fertility decisions, I would have to restrict my sample further by age. Women who were 18 in 2004, when the policy went into effect, are only 34 years old now. I do not include this restriction in my analysis, so my results should be interpreted as an average treatment effect.

Figure 4.4 Shows a plot of college enrollment rates by birth year in the period of study for different states in the US, calculated using the CPS. California is highlighted in red and a thicker line, while the rest of the states are shown in gray. We can see that the overall trend of all states is positive, and California remains in the middle of the pack throughout. Women in California who belong to the 1987 cohort had a college enrollment rate of 0.675. Figure 2.2 shows the estimated college enrollment rates for the 1987 cohort in each state. We can see that the state with highest college enrollment rates is DC with 0.834, and the state with the lowest is Nevada with a rate of 0.517.

2.3.2 Empirical Strategy

The main challenge in evaluating the effects of a policy like this one is that it is hard to find a geographical group that could serve as a reasonable counterfactual to the state of California. The entire US may not be a good counterfactual since college enrollment in

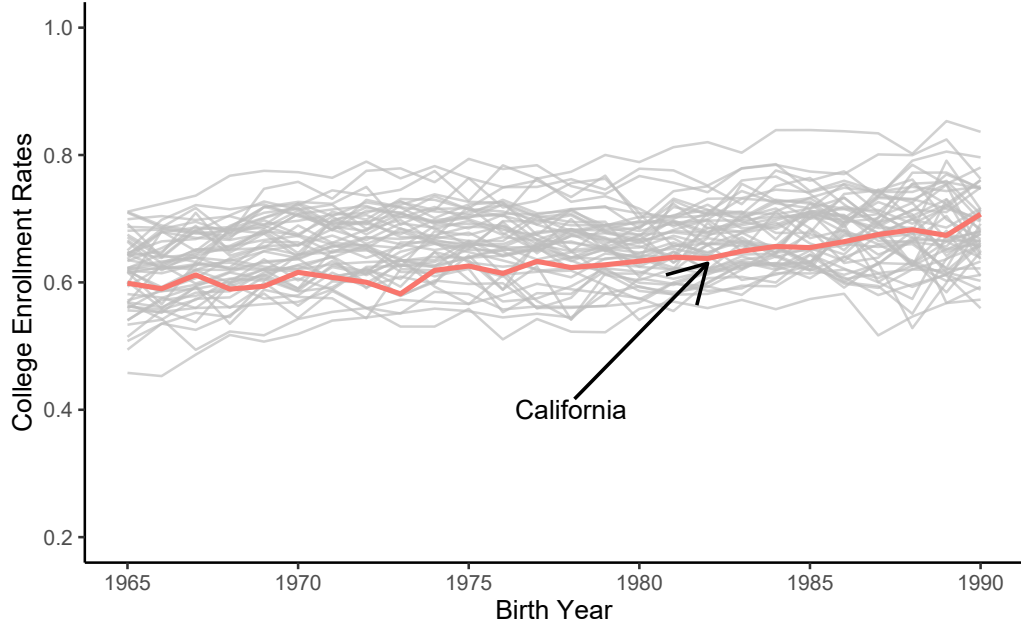


Figure 2.1: College Enrollment Trends in California and Other States

the country as a whole may face different dynamics as the same variable in California. As far as particular states, the choice is not clear: California's neighboring states could be a potential choice with the argument that they are subject to similar economic conditions, but these states are demographically and otherwise different from California. Moreover, using individuals who live close to the state border, between California and Nevada for example, is not a good fit for this research question. While these people may be very similar, endogenous mobility could bias our estimates, especially when one state offers a benefit that the other does not.

The synthetic control method of Abadie et al. (2010) is well suited for the evaluation of aggregated policies such as this. The idea behind this approach is to construct a synthetic California, as a linear combination of untreated states, which will be a good counterfactual for an untreated California. Equation 1 expresses this idea, where \hat{Y}^s is the value of the outcome variable for the synthetic California in time t , ω_i is the weight (between 0 and 1) chosen for each state i , and Y_{it} is the value of the outcome variable in state i , in time t . In this notation, states that are untreated include $i = 1, 2, \dots, N - 1$ and California is indexed by $i = N$.

$$\hat{Y}_t^s = \sum_i^{N-1} \hat{\omega}_i Y_{it} \approx Y_{Nt} \quad (2.1)$$

The method takes as inputs the values of some outcome predictor variables and the

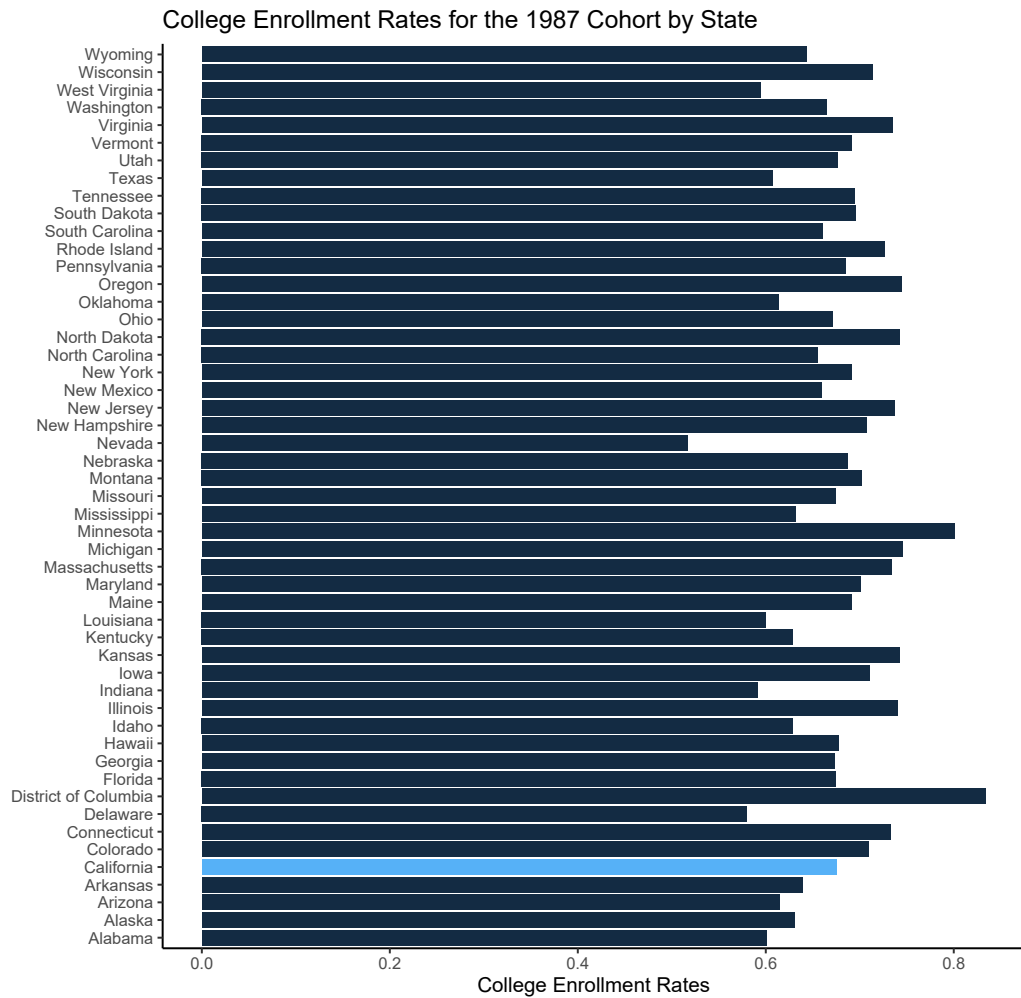


Figure 2.2: College Enrollment Rates in California and Other States for the 1987 Cohort

outcome variable itself for several periods before treatment, and uses them to calculate non-negative weights for every state in the country, such that the weighted average of all states approximates the pre-policy values of the state of California. This is represented in Equation 2, where $\hat{\omega}$ is the vector of all state weights $\hat{\omega}_i$.

$$\hat{\omega} = \underset{\omega}{\operatorname{argmin}} \sum_{t=1}^{T-1} \left(\sum_{i=1}^{N-1} \hat{\omega}_i Y_{it} - Y_{Nt} \right)^2 \quad (2.2)$$

Once the weights are chosen, a synthetic control is then calculated as the weighted average of all states, using the optimal weights. This synthetic control is also projected into the post-policy years, using the same weights. The treatment effect in each period is estimated as the discrepancy between the synthetic control and what actually occurred in California, allowing us to calculate dynamic treatment effects. Equation 3 specifies this.

$$\hat{\tau}_t = Y_{Nt} - \hat{Y}_t^s \quad (2.3)$$

I apply the synthetic control method to my prepared CPS data. I Take college enrollment rates from 1975 to 1985 as pre-treatment observations to fit the synthetic control to. As predictor variables, I use the share of women who report belonging to four racial and ethnic categories, as well as high school graduation rates, and total household income.

I define treated cohorts as those with imputed birth years after 1987. Those born in 1987 were about 17 years old in 2004, when the policy was implemented and thus more likely to still be in high school. College enrollment decisions were more likely to be affected for women who were still in high school at the time of policy implementation. Women who were older and had previously decided not to enroll in college would likely face higher costs if they were to return to school.

2.4 Results

2.4.1 Main Results

Table 2.1 shows the states for which positive weights were chosen for the synthetic control for college enrollment. In this case, only Arizona, Hawaii, New Jersey, New Mexico and Texas were chosen. As Abadie (2019) explains, an attractive property of this method is that it is transparent in how it builds a synthetic counterfactual. Part of this transparency is the fact that the number of states with positive weights is small.

Predictor variables are used to choose state-weights. The relative importance of each predictor variable can be assigned in several ways. In this case, they are chosen proportionally to their variance. These proportions are called v-weights, and they can be found in Table 2.2.

Table 2.3 shows the demographic characteristics of California and its synthetic control. Perhaps unsurprisingly, the hardest control to match is the rate of hispanic women in California. Table 2.2 shows that this predictor variable has a small role in choosing state-weights.

Table 2.1: Synthetic Control Weights: College Enrollment

Weight	State
0.407	Arizona
0.116	Hawaii
0.283	New Jersey
0.059	New Mexico
0.153	Texas

Table 2.2: V Weights

Variable	Weight
Asian	0.021
Black	0.262
Hispanic	0.04
White	0.265
High School Diploma	0
Household Income	0.412

Figure 2.3 shows the trends of real and synthetic California for women with imputed birth years between 1975 and 1990. The dashed vertical line represents the cohort that was first treated. We can see that the synthetic California follows the trend of California quite closely before the policy. The divergence between the two Californias after the vertical dashed line represents the treatment effect. The two lines never cross after treatment, and the size of the gap remains quite constant for the cohorts shown.

The size of the effect may not seem large, but it represents a persistent treatment effect of about 2 percentage points. This is quite large considering the indirect nature of the hypothesized mechanism. After all, the CPFLA only offered about 6 weeks of 60% to 70% of women’s wages, and it would have been difficult to claim for women who were not eligible for FMLA. This effect is also large considering that enrollment rates in California were quite high to begin with, at over 65%. Furthermore, this estimate is large if we consider

Table 2.3: Synthetic Control Balance

	Treated	Synthetic	Sample Mean
Asian	0.136	0.086	0.043
Black	0.060	0.074	0.117
Hispanic	0.418	0.262	0.114
White	0.757	0.750	0.794
High School Diploma	0.856	0.882	0.913
Household Income	67,723.54	67,328.14	61,686.99

it measures effects on human capital investment at the extensive margin. Intensive margin outcomes would likely show larger effects.

2.4.2 Inference

In order to asses significance, I perform a sequence of placebo tests, following Abadie et al. (2010). These placebos perform the same exercise described above, but assign treatment to each one of the untreated states.

The next step to assess statistical significance is to calculate the ratio of pre-treatment RMPSE to post treatment RMSPE for every state, and rank them, following Abadie et al. (2010). The RMSPE ratio is a way to normalize the estimated treatment effect by the pre-treatment fit. The larger the ratio, the less likely the true effect is zero. Ranking all RMSPE ratios allows us to calculate exact p-values. Assessing inference based on these 50 placebo tests is a demanding test of statistical significance.

The histogram of RMSPE ratios is shown in Figure 2.4. As we can see, California’s ratio is larger than all but one state, and it is quite removed from the rest of the ratios in the distribution. The p-value of the effect is calculated as the ranking divided by the number of placebos. In this case, this yields a value of 0.039, which makes my results significant at the 5% level.

A falsification test of sorts is checking for effects in other variables of interest where we would not expect an effect. Figure 2.5 shows the synthetic controls exercise for high school graduation rates. As discussed before, it might be costly for women to adjust some margins of decision on short notice, given previous investments. This is why I did not expect an effect on high school graduation rates, and in fact this placebo test shows no effect. Figure 2.5 shows that the trends of California and its synthetic counterfactual follow each other closely, even after treatment. The high school placebo is also encouraging for the main results, since it provides evidence that the method is able to predict future trends quite well

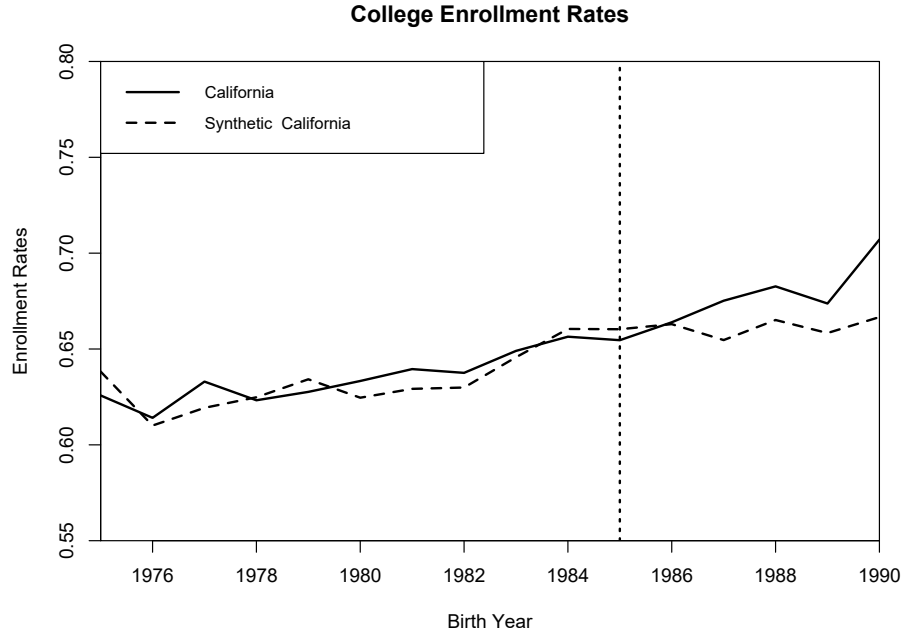


Figure 2.3: College Enrollment Rates for California and Synthetic California

in the absence of treatment.

2.4.3 Heterogeneity by Race and Ethnicity

The average effect found in Figure 2.3 may contain substantial heterogeneity. Given the fact that higher-earning women are more likely to obtain this benefit from their employer, women who expect to be higher-earning are less likely to be affected by the policy. Thus, one can expect heterogeneous treatment effects by some characteristics, such as race and ethnicity, that are related to potential earnings. I therefore construct four strongly balanced panels from the CPS data that contain state-birthyear averages as before, but for each one of the following four categories: asian women, black women, hispanic women, and white women. Hispanic is not mutually exclusive with the other three categories, since hispanic origin is asked as a separate question from race. These categories don't include all women in the CPS, since respondents can select other categories such as Native American, and multi-racial combinations.

Figures 2.6 through 2.9 show the plots for each one of the ethnic and racial categories. Most of them have a reasonable pre-trend fit with the exception of college enrollment of black women, which seems to be very noisy, making it harder to fit by the synthetic control. The only category that seems to show an effect is that of hispanic women. This suggests that the average effect may be higher for hispanic women. However, it is important to note

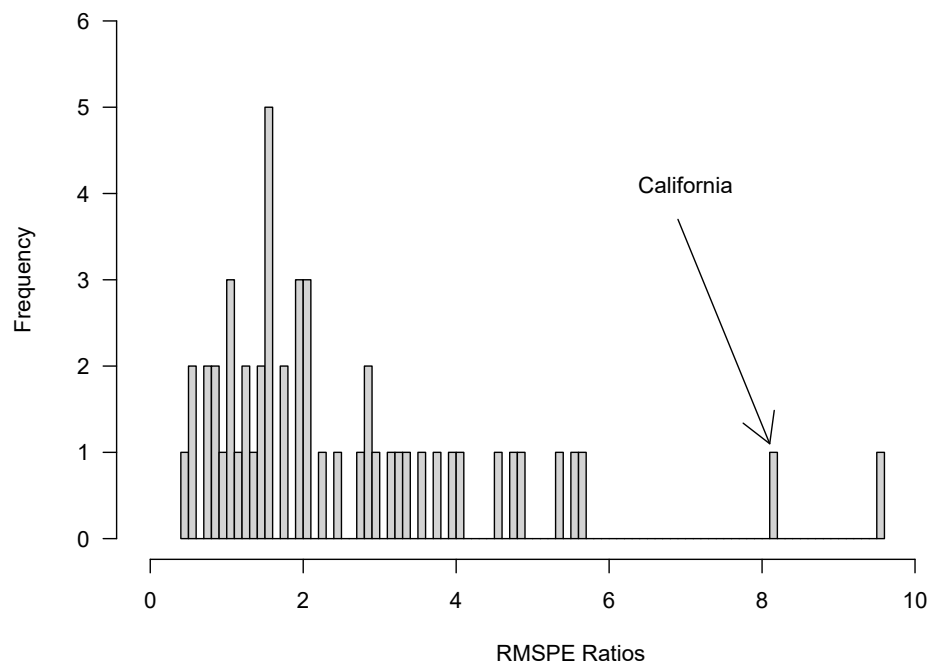


Figure 2.4: Histogram of RMSPE Ratios

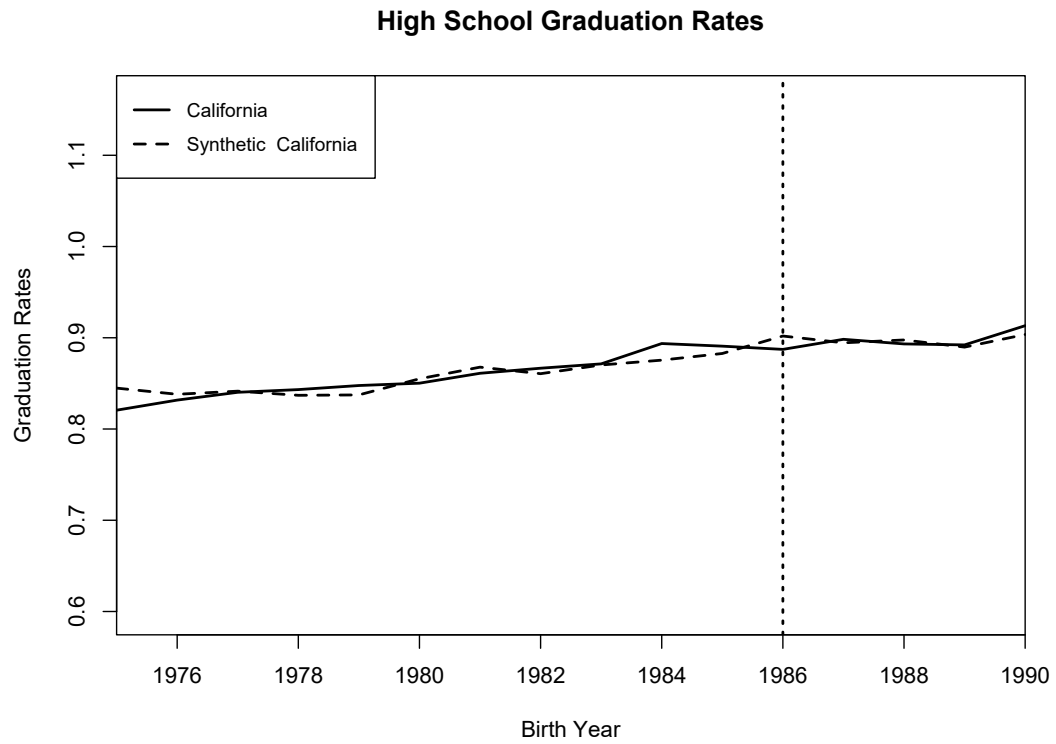


Figure 2.5: High School Graduation Rates for California and Synthetic California

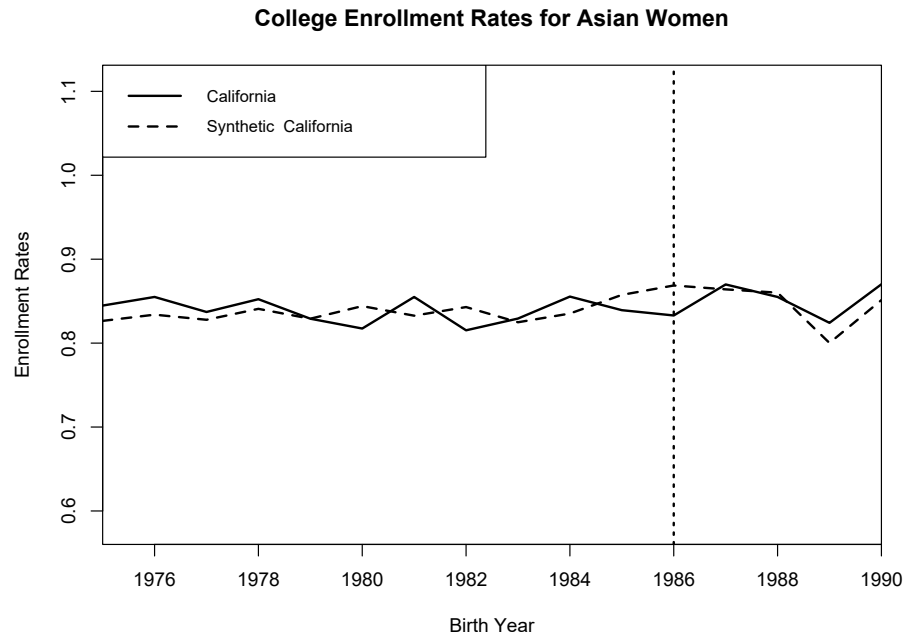


Figure 2.6: Synthetic Control: College Enrollment of Asian Women

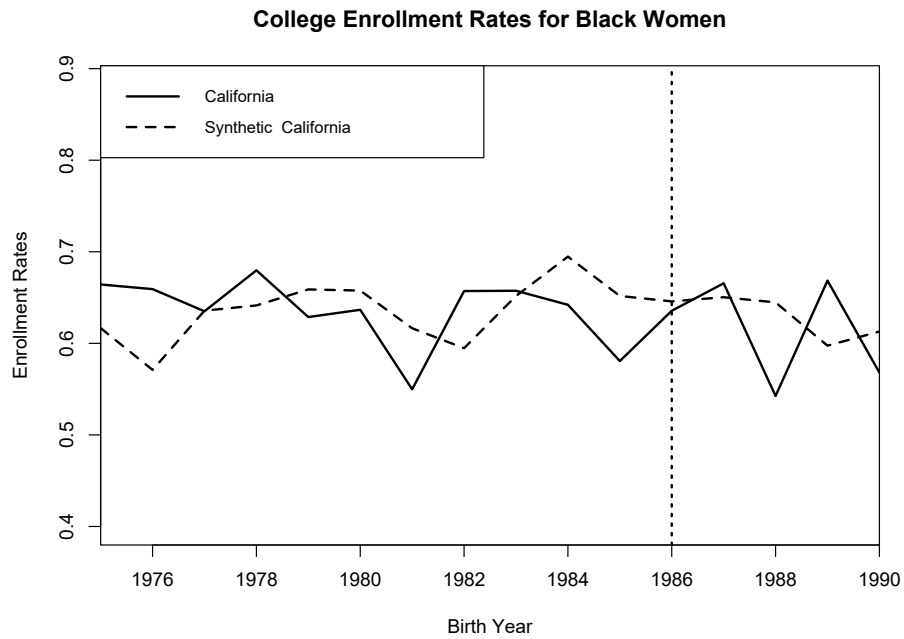


Figure 2.7: Synthetic Control: College Enrollment of Black Women

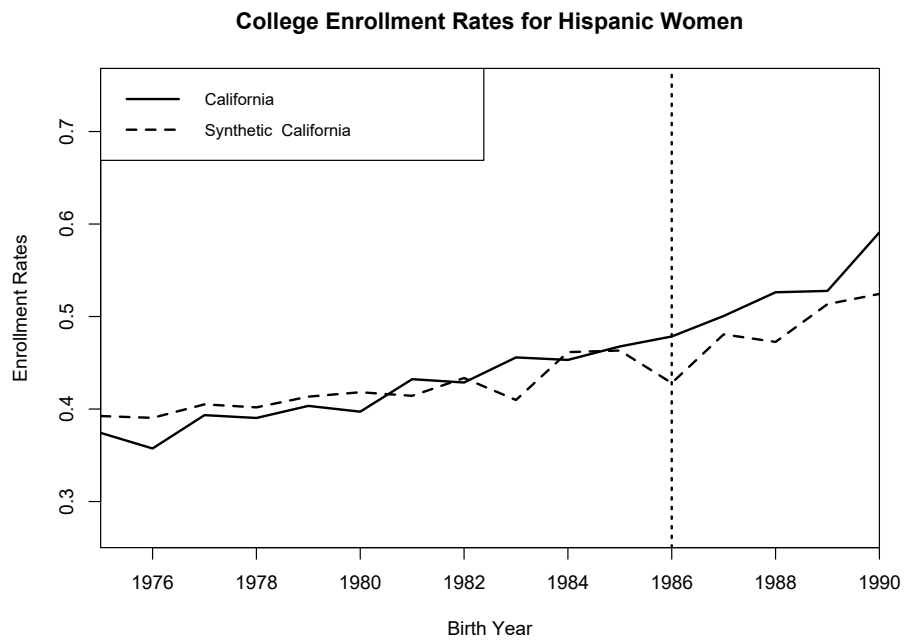


Figure 2.8: Synthetic Control: College Enrollment of Hispanic Women

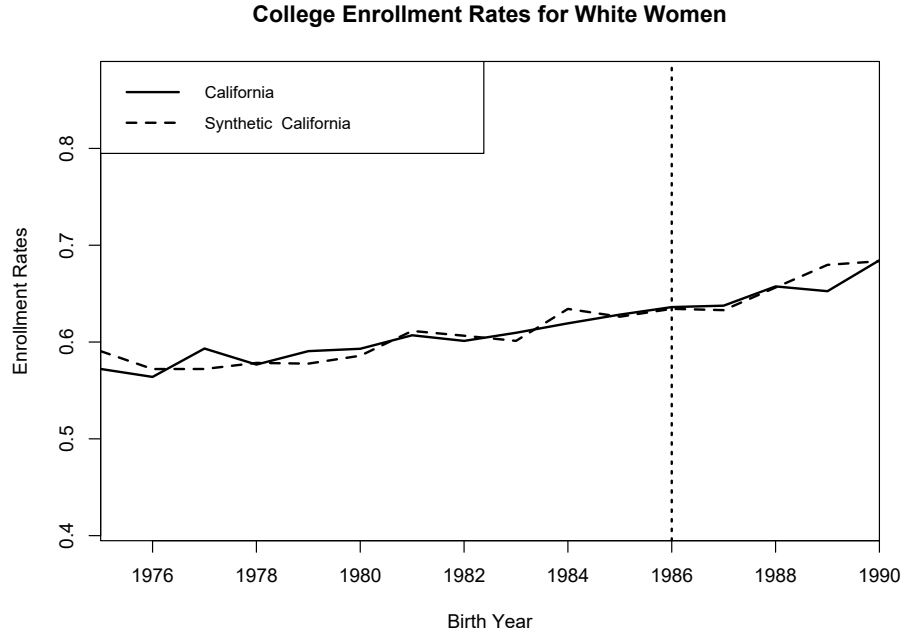


Figure 2.9: Synthetic Control: College Enrollment of White Women

that breaking down the data by race and ethnicity may introduce more noise in the trend, especially in the case of hispanic women outside of California. This could explain why the synthetic control for this group, in Figure 2.8 is not able to match the smooth California trend even before treatment.

The synthetic control method has estimated a statistically significant effect of the introduction of CPFLA on women’s college enrollment. The question remains of where this effect comes from. After all, CPFLA only offered six weeks of partial pay. Only for women at the very margin would this amount of money be enough to tip the balance in favor of a college education. I argue that it is likely that women internalized the implementation of this policy as a reduction in the cost of motherhood on their labor supply. Either because women expected that workplaces had become more parent-friendly or because society as a whole was more accommodating for women’s dual role. To formalize the framework for this mechanism, in the following section I present a simple human capital model.

2.5 Model

Modifying the model from Kuziemko et al. (2018), I characterize women’s optimization problem as follows. Women live two periods, in which they enjoy consumption and leisure, and earn a competitive wage (w) for the labor they sell on the market. They must choose

how much to work on each period, and whether to invest in human capital (schooling) in the first period. If they choose to do so, they will pay tuition (α) on the first period and earn a wage premium (δ , per hour worked) in the second period.

For every hour women choose to work in the second period, they must pay m_i , motherhood costs. Following Kuziemko et al. (2018), I define motherhood costs broadly as any cost mothers have to incur if they work. An obvious example of this is childcare, but m can also include psychological costs, decreases in child quality, and others. The motherhood cost faced by an individual woman comes from an distribution with mean μ . Individual motherhood costs include idiosyncratic variation that represents the fact that although some costs are faced by all, women face different circumstances. Having a network that can provide childcare, living close to childcare facilities, or family members, etc, would be included in ρ_i .

The following is the woman's optimization problem:

$$\begin{aligned}
\max_{c_t, h_t, e} \quad & c_1 - \frac{h_1^{\gamma+1}}{\gamma+1} + \beta \left(c_2 - \frac{h_2^{\gamma+1}}{\gamma+1} \right) \\
\text{s.t.} \quad & c_1 = wh_1 - e\alpha \\
& c_2 = \bar{c} + (\tilde{w} - m_i)h_2 \\
& \tilde{w} = w + \delta e \\
& m_i = \mu + \rho_i \\
& e \in \{0, 1\} \\
& \rho_i \sim \text{iid}(0, 1)
\end{aligned} \tag{2.4}$$

In this model, I assume that women know both components of their own motherhood costs. Because I am looking to rationalize increases in schooling as a result of paid maternity leave policies, and not changes in women's labor supply, I abstract from the issue of uncertainty of motherhood costs. In other words, I am interested in the problem that women are solving when they have to make their schooling decisions, not in the labor outcomes ex-post.

Women can choose not to work, in which case they would not have to pay motherhood costs in the second period, but they would also earn no wages. I incorporate non-labor income in the second period, to account for the income shared by a woman's partner, although this is not the situation faced by all mothers. In this model, mother's childcare labor is unpaid. The model abstracts from several potentially important aspects of motherhood. For example, I abstract from the endogeneity of fertility.

Given a level of education, women can compute the optimal number of hours they will supply on the second period. Hence, this problem is easily solved by backward induction.

There are three possible solutions: 1) women decide not to invest in schooling and not work in the second period, 2) women decide not to invest in schooling and work in the second period, and 3) women decide to invest in schooling and work in the second period. Because there is no uncertainty about the returns to schooling or any other parameter in this model, there is no optimal case in which women decide to invest in schooling but do not work. Table 2.4 shows the value function in these three possible cases.

Even though his model does not include the possibility of the phenomenon documented in Kuziemko et al. (2018) and elsewhere, that women over-invest in education for their labor supply, my results are consistent with the mechanism argued for in Kuziemko et al. (2018): women may be taking cues from society (such as maternity leave policies) as a sign of decreased costs of motherhood, when in fact these costs have increased. In other words, policies like maternity leave could be exacerbating the systematical underestimation of motherhood costs.

Table 2.4: Value Functions

(1) No school, no work in t=2	$\frac{\gamma}{\gamma+1}w^{\frac{\gamma+1}{\gamma}} + \beta\bar{c}$
(2) No school, work in t=2	$\frac{\gamma}{\gamma+1}w^{\frac{\gamma+1}{\gamma}} + \beta[\bar{c} + (w - m_i)^{\frac{\gamma+1}{\gamma}} \frac{\gamma}{\gamma+1}]$
(3) School and work in t=2	$\frac{\gamma}{\gamma+1}w^{\frac{\gamma+1}{\gamma}} - \alpha + \beta[\bar{c} + (w + \delta - m_i)^{\frac{\gamma+1}{\gamma}} \frac{\gamma}{\gamma+1}]$

Given this optimization problem, the question that I ask is how does the choice of solution change with m ? In particular, I am interested in how the schooling choice changes when women are faced with changes in m . Can an increase in the enrollment rate be the product of a decrease in μ ? I begin with some definitions.

Definition 1: Let $\bar{m}_{a,b}$ be the smallest value of m_i that would make a woman indifferent between cases a and b.

We therefore have three different \bar{m} . The value of $\bar{m}_{1,2}$ is of course equal to w , while $\bar{m}_{1,3} = w + \delta - (\frac{\alpha}{\beta} \frac{\gamma+1}{\gamma})^{\frac{\gamma}{1+\gamma}}$. The value of $\bar{m}_{2,3}$ does not have an analytical solution. However, being able to order the magnitudes of these cutoffs is enough to characterize the way in which the schooling choice varies with m .

Proposition 1: Out of the eight plausible orderings of $\bar{m}_{1,2}$, $\bar{m}_{1,3}$, $\bar{m}_{2,3}$, only the following are feasible:

- A: $\bar{m}_{2,3} < \bar{m}_{1,3} < \bar{m}_{1,2}$

- B: $\bar{m}_{1,2} < \bar{m}_{1,3} < \bar{m}_{2,3}$
- C: $\bar{m}_{1,3} < \bar{m}_{1,2} < \bar{m}_{2,3}$

Cases A, B, and C are shown in Figures 2.10 through 2.12. For a given value of m_i , women would choose to be in the case that has the highest value, so the value function is the upper contour of the three graphs presented. For our purposes, only cases A and B are relevant. Case C implies that there is no value of m_i for which schooling would be optimal. In case A, there are values of m_i that would make cases 1, 2, or 3 optimal. In this case, reductions in μ could move women from not working and not acquiring education to just working or to acquire education as well, depending on the magnitude of the change. Reductions in μ could also move some women from case 2 to case 3: women who would have optimized working anyway would now also acquire schooling. I will focus on case A, since it is more general, and because in reality we observe a non-negative share of women in all three cases.

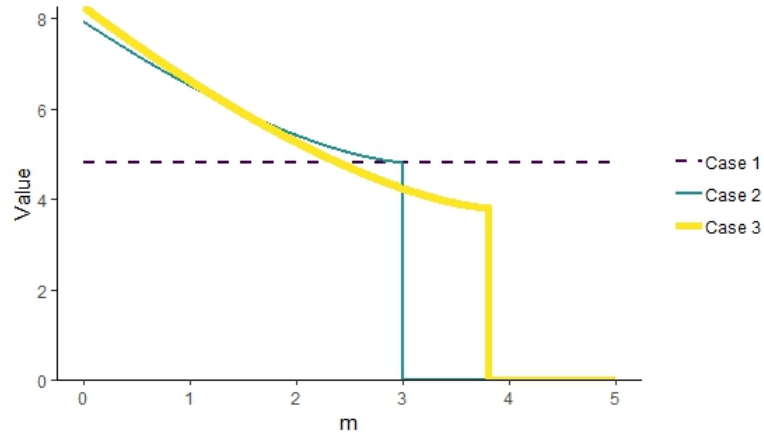


Figure 2.10: Case A

Figure 2.13 shows two distributions of motherhood costs under the case A graph of value functions. Depending on their individual value of ρ_i , women would make different decisions, falling in cases 1 through 3, separated by vertical dashed lines, and labeled in grey. If the mean cost of motherhood decreased, the entire distribution will shift to the left, as illustrated in the bottom panel. This would move some women from case 1 to case 2, and from case 2 to case 3. A larger decrease in mu could even move some women from case 1 to case 3. If we were in the broader case B, where it is never optimal to work without going to school, women would shift from case 3 to case 1. Therefore, a reduction in μ would increase the share of women in case 3, which is consistent with the increase in college enrollment documented in Section 3.

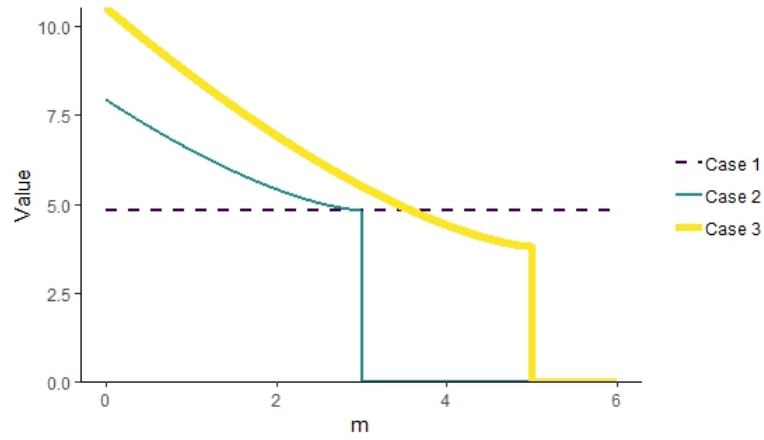


Figure 2.11: Case B

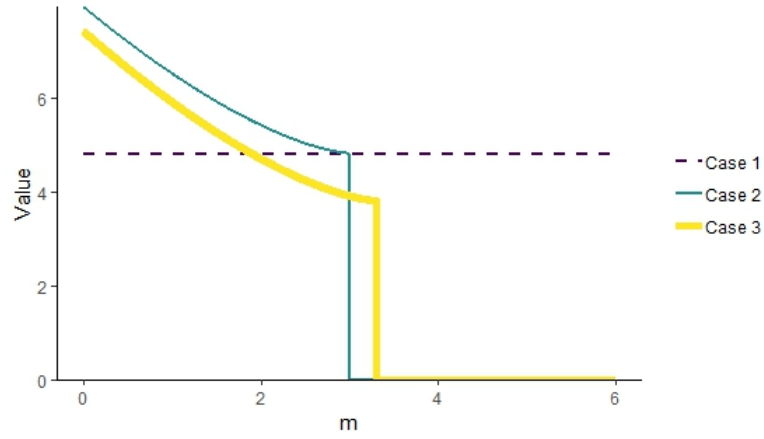


Figure 2.12: Case C

2.6 Other Evidence

2.6.1 Survey Data: General Social Survey

Section 3 estimates a positive effect of the implementation of CPFLA on women's college enrollment. In this section, I present evidence from the General Social Survey (GSS) that suggests the effect occurs through women internalizing the policy as an expected easing of the family-work tradeoffs they face.

The General Social Survey has been collected since 1972, currently on a biannual basis. It contains a large set of questions about opinions held by American adults on topics such as crime, institutions, gender, and others. Its purpose is to inform researchers and policymakers.

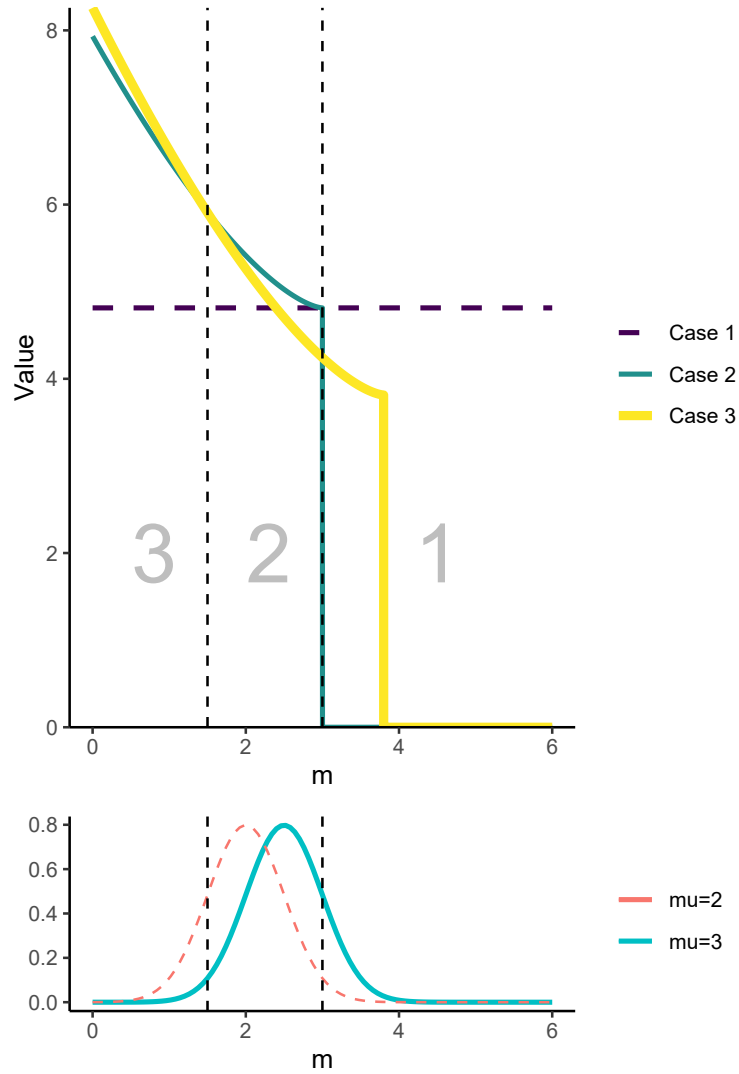


Figure 2.13: Case A With Distribution of Motherhood Costs

For privacy reasons, state markers are not publicly released. I use the regional marker of Pacific, which includes also other states: Washington, Oregon, Alaska and Hawaii. This grouping would mechanically reduce the estimated average effect of the policy with a difference-in-difference estimation. Guided by the heterogeneous effects found in Section 3, and because California has the largest hispanic population of all the states in the Pacific region of the GSS, I run a triple differences specification with the hispanic marker. The triple interaction is composed of: 1) being interviewed before or after the implementation of the policy (2004), 2) living in the Pacific region at the age or 16 or elsewhere in the US, and 3) identifying as hispanic. This triple interaction is more likely to identify the effects of the policy, since most hispanic women in the Pacific region are in the state of California, and

results in Section 3 suggest hispanic women are more likely to be affected within California.

In this specification, treatment is not defined by the year of birth, but rather the year of interview. This is to identify young women’s beliefs before and after the policy came to effect, to see if their beliefs are changed by their awareness of the policy or their exposure to it, regardless of their age. In the sample I include only women who were 25 or younger at the time of interview, to minimize the possibility of the effect motherhood itself or experience in the workplace confound the effect (as in Kuziemko et al. (2018)). Still, I control for whether the respondent has children, as well as some other demographic characteristics.

Table 2.5 shows the results of this specification for some variables related to women’s dual roles as mothers and workers. Column 1 shows the estimated effect of the policy on how rarely women perceive work interfering with family life. Results show that the average effect of the policy on hispanic women is positive and statistically significant. This means that hispanic women in the Pacific region report work interfering with family life to happen less frequently (more rarely) after the policy.

Column 2 finds no effect on how much women agree with the statement that family life suffers as a result of mothers working. This suggests that the motherhood side of the tradeoff was not worsened as a result of the policy. The same holds for Column 3, which estimates the effect on beliefs about preschoolers suffering when their mothers go to work. However, Column 4 estimates that women are more likely to believe that working women can have a warm relationship with their children. This effect is different for hispanic women in the Pacific region than for other women in the region. The total effect estimated for hispanic women is still estimated negative and statistically significant. This result is also consistent with the hypothesis of a perceived decrease on the cost of motherhood that is larger for hispanic women.

Finally, Column 5 estimates a positive difference-in-differences coefficient for disagreeing with the following statement: “A job is alright, but what most women really want is a home and children.” The triple differences coefficient is not statistically significant, albeit also having a positive sign.

Overall, the triple differences estimations presented in this section are consistent with the mechanism rationalized in the model of Section 4. It is plausible that the effect found in Section 3 is driven by a reduction in the perceived cost of motherhood on labor supply. Women, especially hispanic women, as I find evidence that their opinions on topics related to the motherhood-work tradeoff. However, the evidence presented in this section should be interpreted with caution as this specific data set does not allow us to separate women living in California from other women in the Pacific region. Even though the hispanic marker may assist us in this analysis, the structure of the data is not ideal to answer these questions, and results in a small sample.

Table 2.5: Triple Differences Estimation on Variables of Women’s Roles

	(1)	(2)	(3)	(4)	(5)
	How rarely does Work Interfere With Family	Disagree: Family Life Suffers if Mother Works FT	Disagree: Preschoolers Will Suffer Mom Works	Disagree: Working Moms Can Be Warm	Disagree: Most Women Really Want Home and Kids
DDD	2.402*** (0.771)	1.445 (1.655)	-0.494 (1.267)	-2.624** (1.021)	0.151 (1.141)
DD	-0.316 (0.600)	-0.348 (1.203)	-0.055 (0.807)	1.006** (0.460)	1.319*** (0.487)
TE Hispanic	2.185*** (0.593)	1.1222 (1.167)	-0.557 (1.014)	-1.606* (0.881)	1.444 (1.040)
H X A	-1.410*** (0.504)	-0.020 (0.745)	0.581 (0.721)	1.481** (0.734)	0.076 (0.562)
Hispanic N	23	23	23	23	23
N	91	149	148	148	147

Notes: standard errors in parentheses. Controls include number of children, time trend, racial categories, number of adults in the household, mother’s education and having a working mother growing up. Sample includes female participants between 18 and 25 years of age.

2.6.2 Google Trends

When estimating policy effects, it is common to question how much were eligible participants actually aware of the existence of these benefits. If the effect of this policy is indeed through women internalizing a decrease in family-work tradeoffs, it is not necessary that women are aware of the policy details and eligibility rules. A general message being spread would be sufficient, especially if younger women were exposed to it. For example, having more conversations around the topic of maternity leave, and becoming aware of the existence of paid leave could be enough for women to internalize the message.

In order to investigate whether there was interest around maternity leave happening around the time of the policy implementation in California, I use Google Trends data. Google Trends analyzes google web searches and computes an index of relative interest in the specified search terms over a determined period of time. The index takes a value of 100 when the number of searches was the maximum over the specified period of time, and a value of zero when there were none. This index measures interest in the search term over

time, in a specific region.

Google Trends time series begin in January 2004, which means there is not much data on the time before the CPFLA began, in August 2004. However, by looking at the trends starting in January 2004 and comparing different geographical regions, we can get a sense of the relative trends in the conversations or topics that people were reading about.

To eliminate the noise in google trends data, I apply an HP filter to each series. This allows me to extract the underlying trend of the data. This is especially important since, perhaps unsurprisingly, the data is more volatile at the beginning of the series.

Figures 2.14 through 2.16 present the filtered series for different combinations of words related to CPFLA. The solid line represents the HP-filtered index for google searches of that combination of words in the entire United States, while the dashed line is the equivalent but only for searches in the state of California.

In Figure 2.14, we can see that there is a peak in interest in the words “working moms” for the state of California around 2005, close to when the policy began effect (August 2004). The pattern of interest in the United States looks remarkably different. It increases steadily before somewhat flattening around 2015. The difference in trends supports the idea that the topic gained relevance in public discourse in California around the time of the policy implementation. This combination of words is not policy-specific, but rather related to the content of the policy, which suggests that there was public discussion and interest around the topic of women’s dual role as workers and mothers, not just about the specific legislation in question.

Figure 2.15 shows the HP-filtered index for the word combination “Paid Maternity Leave California,” which is more specific to the CPFLA. The patterns of the state of California and the US don’t look that different in this case, but because the word “California” is included in the search criteria, this is not surprising. The fact that interest peaks in 2004 and gradually decreases until after 2010 is encouraging for the hypothesis that policy implementation in fact generated interest in the topic.

Finally, Figure 2.16 shows the HP-filtered index for the words “Paid Maternity Leave”. Patterns differ strikingly between the US and the state of California. These patterns resemble those in 2.14, but are more directly related to CPFLA. What this plot shows is that the pattern of interest in this combination of words in the state of California had a peak around early 2005, decreased and started to gradually increase after around 2011. On the other hand, the entire country had a constant interest at first, and built up interest for this combination of words beginning around 2011. The fact that interest in California peaks around policy implementation, while interest in the rest of the country behaves very differently supports the idea that these issues were being locally discussed.

Specific knowledge of the policy and eligibility rules is not necessary for a more general

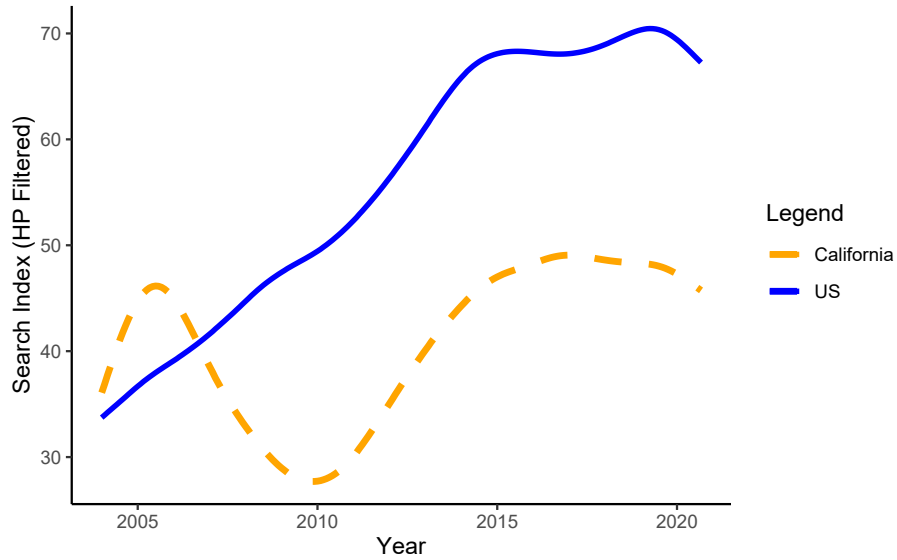


Figure 2.14: Filtered Google Search Index: “Working Mom”

message to seep into attitudes and gain interest in public discussion. The results shown in this section suggest that the policy had some implications in attitudes and public discourse, even if individuals were unaware of the specifics of the policy. Conversations around maternity leave caused by the implementation of the CPFLA may have influenced women’s perceptions of work-motherhood tradeoffs, ultimately modifying their human capital investment decisions.

2.7 Conclusions

I find that the CPFLA induced a statistically significant increase in female college enrollment of roughly 2 percentage points that is persistent. This effect is consistent with women internalizing maternity leave policies as a decrease in the cost of motherhood on their labor supply. This chapter argues that women’s expectation of the effects of a policy are enough to drive behavior change (in particular, schooling choices) even if ex-post effects are not what they anticipated. I provide evidence that interest in these topics was generated around the time when the policy was implemented, and that attitudes around the topics were also affected.

I contribute to the literature of effects of maternity leave policy by studying women’s changes in schooling decisions. Furthermore, my results provide more evidence of the fact that women’s labor supply seems to be constrained by the work-family tradeoffs they face, and their schooling choices respond accordingly. This chapter studies a particular wage

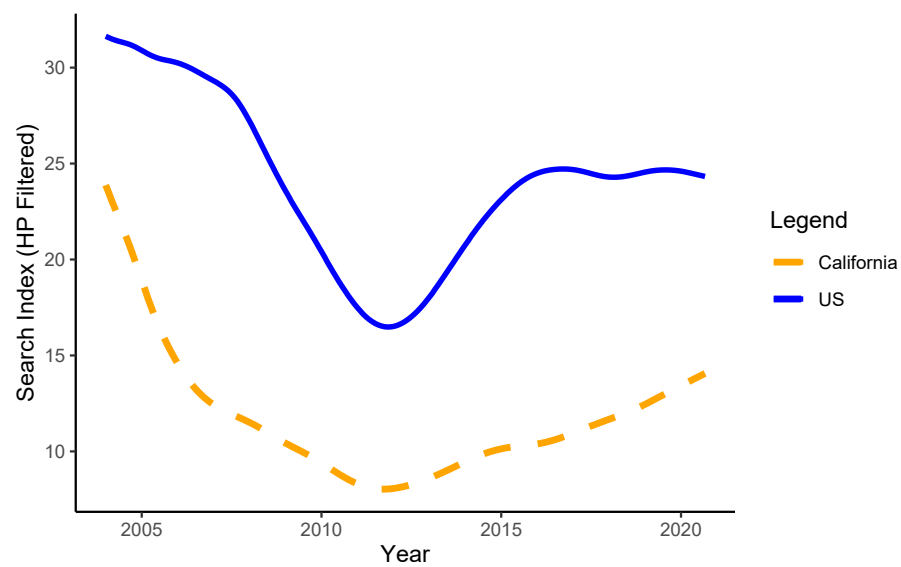


Figure 2.15: Filtered Google Search Index: “Paid Maternity Leave California”

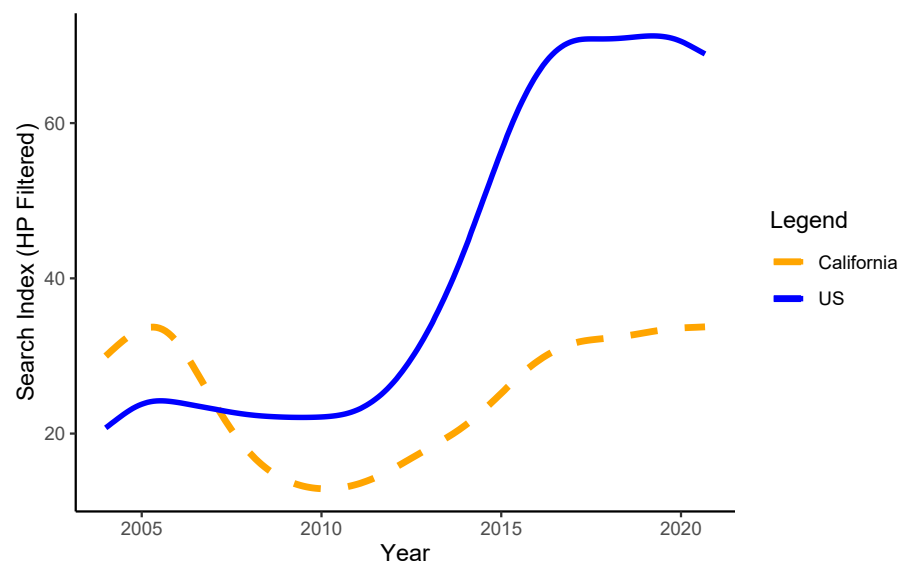


Figure 2.16: Filtered Google Search Index: “Paid Maternity Leave ”

replacement policy. As discussed before, job protection policies were already in place, albeit not accessible to all women. In this sense, the effect found cannot be generalized as an effect of all maternity leave policies. With this data I cannot study how eligibility rules played into the effects I find, but these are potentially important insights, especially for policy makers.

In light of other research finding that the gender gap in labor-market outcomes is stagnant while women's schooling keeps increasing, policymakers and researchers should pay more attention to the broad effects of these policies. Women seem to take maternity leave as a sign of a decrease in the cost of motherhood, and increase schooling in anticipation of higher future labor supply. If paid maternity leave policies don't ease these tradeoffs, or aren't accompanied by such an effect, this may result in women underestimating the cost of motherhood on their labor supply. Statements about the welfare effects of these policies may have to take this into consideration.

More research is needed on the relationship between maternity leave policies, women's expectations of their effects, and the ex-post effects on women's labor-market and related outcomes. In particular, studying the relationship between expectations and the ex-post limiting factors that contribute to persistent gender gaps in labor-market outcomes. It is also important to study how endogenous fertility plays into these mechanisms.

Another choice that these policies could be affecting is that of occupation. Future research should study how these policies may shift women from one occupation (or one college major) to another, and how that could be impacting the gender-wage gap, which has been documented to be related to occupational choices.

Chapter 3

You Know What I Know: Interviewer Knowledge Effects in Subjective Expectation Elicitation

3.1 Introduction

Demographic research has increasingly made use of individuals' subjective expectations about probabilities and the distributions of variables. Such subjective expectations are important drivers of demographic phenomena such as fertility (Delavande, 2008; Mac Dougall et al., 2013; Shapira, 2017) and migration (McKenzie et al., 2013; Shrestha, 2020), and can help us understand their trends and underlying determinants. Furthermore, subjective expectations are related to objective probabilities and can be used to help forecast future trends, for example in the case of mortality rates (Perozek, 2008).

However, the face-to-face surveys commonly used to measure subjective beliefs in developing countries have a potential weakness: respondents' recorded beliefs may be affected by what interviewers know about the phenomenon in question. These surveys sometimes measure subjective beliefs by asking about percent chances directly (Hurd and McGarry, 1995; Lillard and Willis, 2001; McKenzie et al., 2006), but often use visual aids (Attanasio et al., 2005; Delavande and Kohler, 2009; Delavande et al., 2011b) or ask how many of a fixed number of people would have something happen to them (Aguila et al., 2014; de Mel et al., 2008). All three approaches rely heavily on the interviewer to explain the question and encourage the respondent to give a valid answer. These interviewer-subject interactions raise the specter of interviewer effects, and in particular the possibility that interviewer knowledge could inadvertently spill over onto subjects' recorded beliefs.

The effect of interviewer characteristics on survey responses has been documented across

a wide range of contexts. Examples of these characteristics are race and ethnicity (Cotter et al., 1982; Reese et al., 1986; Anderson et al., 1988; Finkel et al., 1991; Davis, 1997; Dionne, 2014; Adida et al., 2016), religion (Blaydes and Gillum, 2013), gender (Becker et al., 1995; McCombie and Anarfi, 2002), and social or cultural proximity (Weinreb, 2006). Respondents may also infer the purpose of the study from interviewers and change their answers as a result, a pattern known as experimenter demand effects (Orne, 1962; Zizzo, 2010; de Quidt et al., 2018). This body of research shows the importance of social interactions in the interview setting for recorded survey responses, and how interviewer characteristics may impact this interaction. An extensive literature has also explored the methodology of subjective belief elicitation (Delavande, 2014). However, to our knowledge, no previous paper has studied the role of interviewer knowledge in driving survey responses.

Leveraging a randomized experiment that used interviewers to implement an information treatment, we show that interviewer knowledge has an effect on respondents’ recorded beliefs. The experiment was designed to investigate how information about the true transmission rate of HIV affects risk-taking (Kerwin, 2020). Interviewers were taught the true HIV transmission rate mid-way through the baseline survey, in order to conduct the information treatments for the study. The study respondents were randomly divided into control surveys, which happened before the interviewers learned the information, and treatment surveys, which happened afterwards. We use data from the baseline surveys, when treatment-group respondents had not yet been taught the risk information themselves, but were interviewed by people who had been taught it.

Interviewer knowledge matters for recorded risk perceptions. Comparing the baseline surveys across study arms, we find that interviewers who were exposed to the information treatment elicit lower HIV transmission rate perceptions from respondents. Reported beliefs are significantly shifted by the interviewers’ knowledge, decreasing by about nine percentage points, or roughly 0.3 SD of the control-group belief distribution. This finding can help explain the puzzling finding that people’s preferences and beliefs appear to be very unstable in panel surveys (Chuang and Schechter, 2015; Mueller et al., 2019). If recorded responses are heavily shaped by interviewers’ knowledge and beliefs, then people’s answers may appear to be much more unstable than they really are.

In addition to shedding light on the role of interviewer knowledge in driving survey responses, our study also builds on the previous literature on interviewer effects by isolating the causal effect of a specific interviewer characteristic—knowledge. Past studies of interviewer effects have been able to exploit the exogenous assignment of interviewers to respondents, but have been limited by the fact that the interviewer characteristics in question are both fixed and correlated with other attributes. For example, race is correlated with income and socioeconomic status, and a wide range of interviewer characteristics can

all affect responses simultaneously (Di Maio and Fiala, 2019). Because interviewers in our study were exogenously shocked with new information about HIV transmission rates, we can isolate the causal effect of knowledge alone. This is the first study we are aware of that has been able to identify the causal effect of a single interviewer characteristic. This is possible because knowledge, unlike the other characteristics that are typically studied, is malleable: it can be changed quickly, whereas even many non-fixed traits like education levels can be changed only slowly, and others such as age cannot be changed at all.

We can identify several channels through which interviewers' knowledge affects recorded risk perceptions. First, interviewers who underwent the training primed respondents to give answers that match the exact training content. The training explained that the annual transmission rate of HIV between an HIV-positive spouse and an HIV-negative spouse who have regular unprotected sex is 10%. Consistent with a priming story, treatment-group respondents are 4.3 percentage points more likely to (incorrectly) report that the per-act probability of HIV transmission is exactly 10%.

A related mechanism by which interviewer knowledge affects recorded risk perceptions is through nudging respondents to give lower answers. Evidence for this comes from an aspect of the survey design: if a respondent answered exactly 50% for any risk perception question, interviewers were taught to follow up and see if they were simply unsure; if so, they were asked for their best guess, following Hudomiet et al. (2011). Interviewers who underwent the training were less likely to elicit higher numbers when they ask respondents to provide a best guess in this situation. This suggests that interviewers who have been exposed to the information treatment are nudging participants away from higher answers. The same pattern could also affect the initial responses to the questions.

The strength of respondents' priors may affect how much interviewer knowledge matters. The effect of interviewer training is smaller for more educated respondents, and falls to zero for respondents who reached at least Form 2 (10th grade) in school. This may be due to the fact that students in Malawi learn about HIV transmission during Form 2, and are exposed to a narrative that claims HIV is highly contagious. While the information taught during Form 2 diffuses through the population as a whole, more-educated respondents are exposed to it directly, and thus likely feel more certain about their beliefs. This makes them less susceptible to the interviewer's nudges to report lower risk beliefs.

We suggest several ways to correct for interviewer knowledge effects. Interviewer recruitment for face-to-face surveys should try to match the population of respondents, and interviewer training should emphasize the possibility of unintentional spillovers and the need to treat all respondents consistently. When designing information experiments, researchers should consider running baseline surveys simultaneously across groups or separating the information treatment from surveys, although these approaches have their own drawbacks.

Even when no information is provided, knowledge spillovers remain likely—interviewers can vary widely in their knowledge and beliefs, and so teaching them about topics relevant to the survey may improve data quality. One promising avenue is to eliminate the interaction between interviewer and respondent by performing surveys via audio computer assisted self interviewing (ACASI) which can work even in contexts with low literacy and numeracy. However, ACASI may lead to lower data quality than face-to-face interviews; further work in this area would be invaluable.

The remainder of this paper proceeds as follows. Section 1 reviews the role of subjective beliefs in demography, provides background on Malawi’s HIV epidemic, and describes the previous literature on subjective expectations about HIV in Malawi. Section 2 describes the experiment and the data, and Section 3 presents the empirical strategy we use to evaluate interviewer knowledge effects. Section 4 shows how interviewer knowledge affects measurements of respondents’ subjective expectations. In Section 5 we explore mechanisms for the effects, and in Section 6 we discuss potential corrections for interviewer knowledge effects. Section 7 concludes.

3.2 Background

3.2.1 Subjective Expectations and Demographic Decisions

Subjective expectations play a key role in driving demographic patterns and people’s responses to them. They drive contraception choices and fertility (Delavande, 2008; Mac Dougall et al., 2013; Shapira, 2017) as well as migration (McKenzie et al., 2013; Shrestha, 2020). Perceived mortality risks affect whether people engage in life-threatening activities—see e.g. Delavande and Kohler (2016) for HIV, León and Miguel (2017) for transportation choices, and Benneer et al. (2013) and Keskin et al. (2016) for water safety—and thus actual mortality rates. The effects of subjective expectations often spill over between demographic choices and other domains. For example, women’s education and career decisions depend on their beliefs about the costs of raising children—which can differ sharply from reality (Kuziemko et al., 2018). Similarly, women tend to systematically underestimate their fecundity at young ages and overestimate it at older ages (Mahony, 2011).¹

¹In addition to uncertainty about their material circumstances, people may also be uncertain about their own futures or the correct course of action; periods of uncertainty can cause intersections between aspects of people’s lives that usually seem separate, such as work and romantic relationships (Johnson-Hanks, 2017). This uncertainty about the correct course of action is itself shaped by uncertainty about material facts, such as the risks of mortality and miscarriage (Trinitapoli and Yeatman, 2018).

Data on people’s subjective expectations are also a potentially-useful tool for demography research. Subjective mortality probabilities may be useful predictors of actual mortality rates (Perozek, 2008) and are correlated with known predictors of mortality (Delavande et al., 2017), although there is also evidence that people’s subjective mortality beliefs have systematic biases (Elder, 2013). Despite their limitations, subjective mortality beliefs may still be valuable: people form them using risk factors such as parental health and longevity that objective mortality rates cannot account for, and they affect risk-taking behaviors (Dormont et al., 2018). Similarly, self-rated health is a useful predictor of mortality (Burstrom and Fredlund, 2001). Subjective beliefs about health can help forecast future mortality rates: they are available earlier than objective predictors of mortality, and they predict mortality even conditional on objective measures of health status (Idler and Benyamini, 1997).

There is a crucial distinction between individuals’ subjective expectations about risks and other variables and the true values of these figures. Much social science research assumes that people know the true values of numbers, but recent research has focused on measuring people’s actual perceptions, which can be quite different from the truth (Manski, 2004). Consider the case of subjective expectations about mortality rates. These can differ from true population-average mortality rates in three key ways (Delavande and Rohwedder, 2011). First, they measure a variable that has not yet been observed because the population answering the survey questions is still alive. Second, they may be measured with error relative to the person’s true beliefs. Third, they reflect individuals’ beliefs about what will happen, rather than the truth.

3.2.2 The HIV Epidemic in Malawi

Malawi has been dealing with a severe HIV epidemic for several decades and the disease has had major effects on its population. The prevalence of the virus has been stable at around 10% of the population for roughly the past decade (National Statistical Office - NSO/Malawi and ICF, 2017).² The expansion of access to antiretroviral treatment (ART) for HIV has drastically improved the situation for HIV-positive people in recent years. Starting in 2016, Malawi implemented a universal test-and-treat policy, so that all HIV-positive people had access to ART (Alhaj et al., 2019). Testing rates are still low for men, but most women get access to treatment because there is strong pressure to accept the nominally-voluntary HIV tests during antenatal care visits (Angotti et al., 2011).³ Even

²A small apparent drop (to a prevalence of 9%) in the 2015 DHS was the result of a change in methodology; measured on a consistent basis, the prevalence was essentially unchanged from the 2011 survey.

³This is part of a systematic pattern of HIV prevention efforts targeting women and excluding men (Watkins, 2011).

with the expansion of access to treatment, however, HIV is still a major issue in people's lives.

The large scale of Malawi's HIV epidemic has led to extensive research by social scientists on how it impacts people's lives. Most prominently, this includes the Malawi Longitudinal Study of Families and Health (MLSFH), which has been collecting demographic, socioeconomic, and health information on the same households since 1998. The MLSFH also embeds a novel ethnographic study, the Malawi Journals Project (MJP), in which Malawians record everyday conversations about HIV/AIDS.

3.2.3 Subjective Expectations About HIV in Malawi

An important focus of research on HIV in Malawi has been on measuring people's subjective beliefs about the disease and understanding how those beliefs affect their behavior. The MLSFH measures both people's perceptions about HIV and their sexual activity, and has an embedded experiment in which respondents were incentivized to learn their HIV status (Fedor et al., 2015; Thornton, 2008). It was also used as a platform to design and study an innovative technique for capturing subjective probabilities using visual aids (Delavande and Kohler, 2009). This work was an important contribution to a literature that shows that eliciting subjective probability beliefs is feasible in low and middle-income settings (Attanasio, 2009; Delavande, 2014; Delavande et al., 2011a).

A core finding of the work on subjective beliefs about HIV in Malawi is that people substantially over-estimate their likelihood of being HIV-positive (Bignami-Van Assche et al., 2007; Anglewicz and Kohler, 2009). Relatedly, they also over-estimate the transmission rate of the virus by several orders of magnitude (Delavande and Kohler, 2016; Kerwin, 2020). Extensive research has tried to understand how people form these beliefs. One channel is through HIV testing: Malawians who learn they are HIV-positive lower their beliefs about the transmission rate of the virus (Delavande and Kohler, 2012); this may be because they realize they have not yet transmitted the virus to their sex partners. Qualitative evidence from the MJP supports this quantitative finding. Kaler and Watkins (2010) found that people are ambivalent about testing: they think that it will always lead to a positive result, followed by death. They also found that as a result of thinking that HIV tests mostly turn out positive, people overestimate the transmission rate of HIV.

Information from HIV tests spreads beyond the person being tested. Spouses typically tell each other about their HIV test results, although HIV-positive women are less likely to share their status (Anglewicz and Chintsanya, 2011). More broadly, subjective expectations about HIV risks spread through social networks (Helleringer and Kohler, 2005; Kohler et al., 2007). People also draw inferences about HIV risks from their own experiences.

For example, when young women marry, they are more likely to think they are at risk of contracting HIV in the future—possibly because they know or suspect their husbands are unfaithful (Grant and Soler-Hampejsek, 2014).

Another line of research has shown that subjective expectations about HIV matter: people respond to their perceived risks of being HIV-positive. It is perceived, rather than actual, HIV status that drives condom use for women, even when one’s actual status is known (Anglewicz and Clark, 2013). Ethnographic evidence from the MJP shows a similar pattern for men. They assume they are HIV-positive even with no medical evaluation or signs of AIDS, which drives further risky behavior (Kaler, 2003; Kaler et al., 2015). Many people are uncertain about their HIV status, and this uncertainty affects their fertility intentions (Trinitapoli and Yeatman, 2011).

In addition to changing their behavior in response to their perceived HIV status, people also respond to their perceived chance of contracting the disease. (Grant and Soler-Hampejsek, 2014) show that women may use divorce to protect themselves if they believe they are at high risk of contracting HIV, mirroring the finding by Anglewicz and Reniers (2014) that HIV-positive people have higher rates of widowhood and divorce. Women who anticipate that they will contract HIV in the future invest more in their children’s education (Grant, 2008). The causal effect of risk perceptions on behavior also holds for probabilistic beliefs of the kind studied in this paper (Delavande and Kohler, 2016; Kerwin, 2020).

3.3 Data and Empirical Design

We use data from an experiment designed to study the effects of risk perceptions on risk-taking behavior (Kerwin, 2020) that was conducted in the Zomba District of Malawi from August to December 2012. The experiment randomly assigned one-half of respondents (stratified by distance to the trading center and gender) to receive information about HIV transmission risks at the end of the baseline survey. Treatment-group participants were read an information script that explained the actual HIV transmission rate for couples with one infected partner that have regular unprotected sex (about 100 times per year, on average). The true transmission rate is 10% per year (Wawer et al., 2005), far below what Malawians typically believe. In our sample, the average risk belief is about 90% per year, and nearly one-half of our sample thinks that the transmission rate from just a single exposure is 100%.

The risk information was provided by the survey interviewers, using a script and a set of visual aids that were built into the treatment-group surveys. The interviewers themselves were taught the risk information and how to conduct that survey module via a two-day training session that took place halfway through the baseline data collection. All the control-group surveys were scheduled to occur before this training session to minimize the risk of

contaminating the control villages, following Godlonton et al. (2015).⁴

The interviewers seem to have been unaware of the actual HIV transmission rate prior to the training session, and thus it likely strongly shifted their beliefs about HIV risks. Although we lack direct data on their beliefs prior to the information session, two sources of evidence support this claim. First, the interviewers all lived in or close to the study area, so the pre-training data for the control group is a reasonable proxy for their beliefs. Less than 2% of our control group thought the annual risk of HIV transmission was below 20% at baseline. A second piece of evidence comes from the training session itself. The interviewers expressed surprise at the information they were taught, and many were initially reluctant to believe it. To help convince them, project staff had to show them the original research study (Wawer et al., 2005) as well as the section of the Malawi National AIDS Commission website that listed the HIV transmission rate.

The fact that the interviewers were taught new information just before they started to survey the randomly assigned treatment group allows us to study how that information affected survey responses. We use the interviewer training session as a treatment and study how that changes the recorded baseline beliefs of respondents. Comparing the baseline beliefs between the treatment and control groups allows us to estimate the effect of interviewer knowledge on recorded risk beliefs.

Our principal outcome measure is respondents' recorded subjective risk beliefs on the baseline surveys. This variable was collected by asking respondents to estimate proportions of a fixed number of people—for example, “If 100 men, who do not have HIV, each have sex with a woman who is HIV-positive tonight and do not use a condom, how many of them do you think will have HIV after the night?” The questions cover transmission rates per act and per year for both protected and unprotected sex. Respondents picked integers between 0 and 100 in response to each question.¹noteFor the per-year question about unprotected sex, the correct answer is 10; for the per-act equivalent, the closest possible answer to the truth is 0. In the absence of condom failures, the correct answer for both the per-year and per-act condom-protected sex questions is 0. This style of question to assess expectations has also been tested and validated by previous research in Malawi (Chinkhumba et al., 2014; Godlonton et al., 2015; Kerwin et al., 2011).⁵ Interviewers had no incentive to record specific answers to this question but instead were incentivized to record answers accurately:

⁴There were a handful of control surveys that did take place after the training session. These were cleanup surveys that happened because selected respondents were not available at the time of the scheduled baseline interview, and had to be interviewed afterwards.

⁵These questions measure the respondents' perceived risk of contracting HIV from various sexual behaviors, not their perceived probability of currently being HIV positive. Our measured probabilities are comparable to other measurements from Malawi. For example, Delavande and Kohler (2012) found a mean per-act risk for unprotected sex of 87%, compared with 83% in our data.

random back-checks were used to check that surveys were actually conducted and responses were recorded correctly.

Our sample of respondents includes 1,292 individuals from 70 villages who have both valid baseline and endline survey data. The two study arms were balanced on observable exogenous variables. The experiment we use was not designed to study the interviewers, and so we have very limited data on their characteristics based on administrative records. Of the 14 interviewers in our sample, half are female and half male. The even gender split was chosen intentionally to facilitate gender-matched interviews: all male respondents were interviewed by men, and all women were interviewed by women. All interviewers had completed secondary school (a screening requirement imposed during hiring), and most had graduated recently (and thus they were in their 20s). They were recruited from the local area but were not assigned to survey anyone they knew personally.

3.4 Empirical Strategy

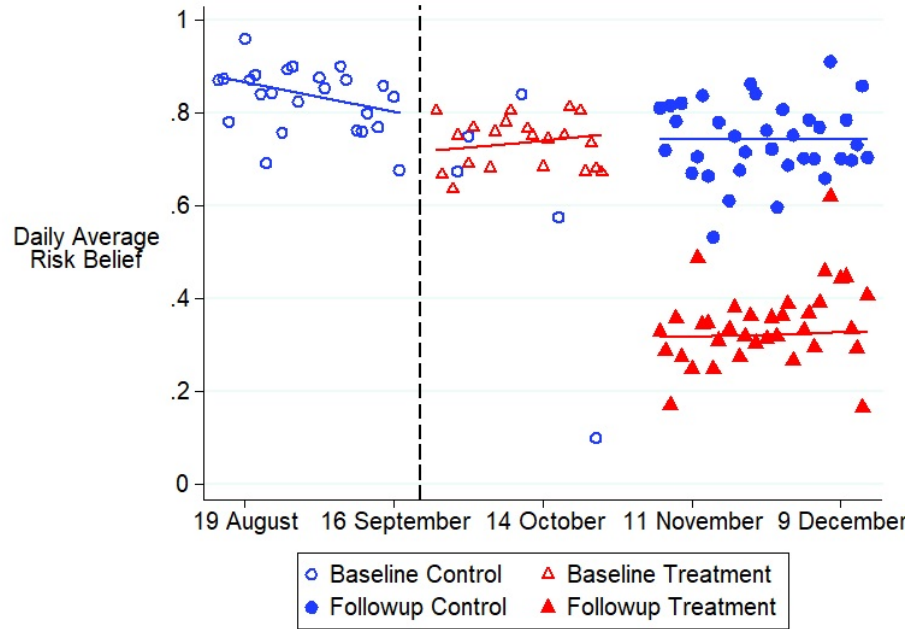
To study the effect of interviewer knowledge on respondents' recorded risk beliefs, we compare the recorded baseline beliefs of the treatment and control groups. Our main regression specification is Eq. (3.1), where Y_i is either a measure of risk beliefs at baseline or an indicator variable for specific values of the risk belief at baseline. The dummy variable T_i takes a value of 1 for respondents in the treatment group and 0 otherwise. Our treatment is thus defined as having been interviewed at baseline by a more-knowledgeable interviewer. We control for sampling strata fixed effects, Z_i , and interviewer fixed effects, I_i ; the latter allow us to rule out the possibility that interviewer characteristics other than knowledge are driving our results. We also control for W_i , a sexual activity index based on the first five variables in the balance table (see the section Alternative Explanations for further discussion). All standard errors are adjusted for clustering by village.

$$Y_i = \alpha + \beta T_i + Z_i' \eta + I_i' \gamma + \delta W_i + \epsilon_i. \quad (3.1)$$

To understand the mechanisms behind the effects, we interact the treatment indicator with respondent characteristics (Eq. (3.2)). We de-mean all covariates before interacting them with the treatment indicator, so the main effect of the treatment can still be interpreted as the sample-average treatment effect (Imbens and Rubin, 2015, p. 247).

$$Y_i = \alpha + \beta T_i + \gamma T_i \times X_i + \delta X_i + Z_i' \eta + I_i' \gamma + \delta W_i + \epsilon_i. \quad (3.2)$$

Figure 3.1: Measured Risk Beliefs over Time, by Study Arm



Notes: Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Risk beliefs are the perceived probability of contracting HIV from a single unprotected sex act with an infected partner. Each point represents the mean value of the risk beliefs for a given day; baseline control beliefs are hollow circles, endline control beliefs are solid circles, baseline treatment beliefs are hollow triangles, and endline treatment beliefs are solid triangles. The lines are linear fits of beliefs on date for a given date range and study arm. The dashed vertical line indicates the date of the training sessions when the survey interviewers were trained to provide the information treatment about HIV transmission risks.

3.5 Results

3.5.1 Main Estimates

Interviewers exposed to the information treatment elicit lower risk perceptions. Figure 3.1 shows the daily average recorded risk beliefs for the treatment and control groups over time. The first group of observations represents the control-group beliefs at baseline, when neither the interviewers nor the respondents knew the content of the information treatment. After those surveys were complete, the interviewers learned the content of the information treatment (vertical dashed line) and then conducted the baseline treatment surveys. We can see that risk beliefs of the treatment group are lower than those of the control group.

There are five days with control-group baseline data after the information treatment. These are cleanup baseline surveys for the control group that were conducted after the bulk of the baseline control-group surveys were finished and that took place after the interviewer training session. This happened when respondents were not available at the initially scheduled baseline interview. The distribution of beliefs for these observations is closer to that of the treatment group than to the rest of the control group, lending support to the idea that interviewer knowledge specifically—and not some other factor that is imbalanced across study arms—is causing the mean difference between baseline treatment and control recorded beliefs.

Further support for the idea that the change in beliefs is due to interviewer knowledge is evident in the endline risk beliefs. First, the endline risk beliefs allow us to reject the possibility that the treatment group simply accidentally received the risk information prior to answering their baseline survey questions. The direct information treatment effect on risk beliefs (the gap between the endline risk beliefs for the treatment and the control groups) is much larger than the treatment-control difference at baseline.

Second, the control-group endline beliefs are very similar to the treatment-group baseline beliefs, which is completely consistent with a model in which recorded beliefs are moved by interviewer knowledge. Neither the treatment group at baseline nor the control group at endline had been directly told the information about HIV transmission risks, but both were interviewed by interviewers who did know the information. As a result, both sets of risk beliefs are shifted downward relative to the control-group baseline beliefs, and they also have similar average values to one another.

Table 3.1 presents our main results numerically. Each column represents a measure of a different HIV transmission risk: measured per act or per year, and when using condoms or having unprotected sex. For all four measured risk beliefs, the coefficient on the treatment (interviewer training) is negative and significant. In the case of the per-act transmission risk for unprotected sex, the coefficient is 9.3 percentage points, or about 0.3 standard deviations. The magnitude of the effect is relatively large, especially considering that it is an unintentional spillover: respondents were not directly exposed to the information treatment. As shown in Figure 3.1, the effect at endline, when participants themselves were exposed to the information treatment, was larger: 38.4 percentage points for the perceived per-act transmission risk for unprotected sex.

Participants in the control group had average risk beliefs that were substantially larger than the true risk of HIV transmission in each one of those cases. For example, the true value of the per-year transmission rate for unprotected sex is about 10% (Wawer et al., 2005), but the average respondent in the control group thought the risk was 83%, and well over one-half of respondents thought the risk was 100%. Baseline beliefs for the control

Table 3.1: Effects of interviewer knowledge on reported risk beliefs

	(1)	(2)	(3)	(4)
	Outcome: HIV transmission risk belief			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment	-0.0745*** (0.0191)	-0.0352*** (0.0128)	-0.0366*** (0.0128)	-0.0850*** (0.0157)
Control-group mean	0.827	0.927	0.123	0.236
Control-group SD	0.264	0.169	0.218	0.279
Observations	1,282	1,277	1,284	1,277

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

group have the correct ordering in terms of which risk is higher, but all the average levels are higher than the true infection risks.

Interviewer training decreased recorded risk beliefs for all four measures, even though the training discussed only the per-year risk for unprotected sex, shown in column 2 of Table 3.1. Column 1 shows an effect of 9.3 percentage points (0.35 SD); columns 2 and 4 show effects of 4.8 and 7.9 percentage points, respectively (0.28 SD each). Column 3 shows the smallest effect, at 2.7 percentage points (0.12 SD), corresponding to the per-act, condom-protected transmission risk.⁶ This variable has the lowest control-group mean overall, so a smaller effect is not surprising, and we can still reject a zero treatment effect. Moreover, the risks of condom-protected sex are simply scaled-down versions of the risks from unprotected sex, so changes in those variables should be smaller.

The fact that interviewer knowledge changes responses for risk beliefs that were not explicitly targeted suggests that interviewers internalized the information and actually changed their beliefs about transmission risk, as opposed to memorizing the one figure that was presented to them. Interviewers knew that the four measures of transmission risk were related, and when they adjusted their beliefs for one, it affected their beliefs for all the others. Thus the threat of interviewer knowledge effects appears to be quite general:

⁶The results in Table 1 are qualitatively identical if we omit the sexual activity index or the interviewer fixed effects from the controls.

knowledge spillovers occur not only with directly provided information but also with the implications of the information.

3.5.2 Alternative Explanations

Interviewer Experience

Another potential explanation for our findings is interviewer experience. The trajectory of the pretreatment trend in the first portion of Figure 3.1, if extended, would intersect the level of recorded risk beliefs in the second portion. This could have happened if interviewers improved over time at asking the relatively complicated questions on the subjective expectation module. Our basic results in Table 1 do not rule out the possibility that the estimated treatment effects are due to interviewer experience alone.

To examine this possibility in further detail, we present a set of regression discontinuity (RD) plots in Figure 3.2. These graphs are produced using the (`Calonico et al., 2015`) `rdplot` Stata command to automatically bin the data and fit polynomial curves on either side of the discontinuity. The binned averages are shown using dark gray dots, with black lines for the polynomial fits. The light gray regions show 95% confidence intervals for the bin-specific averages.

We also show the estimated treatment effects from RD models in Table 3.2, using the `rdrobust` Stata command (`Calonico et al., 2017`). This command automatically selects data-driven bandwidths and computes robust bias-corrected p values (`Calonico et al., 2014`).

Panel a of Figure 3.2 shows the main comparison of interest: before versus after the training session, per-act HIV transmission risk beliefs for unprotected sex. Two conclusions are clear from the graph. First, the steep downward trend apparent in the first portion of Figure 3.1 is partly an artifact of fitting a linear trend to daily average risk beliefs as opposed to the underlying survey data. Fitting a flexible polynomial to the actual survey responses reveals a slight downward trend to the left of the discontinuity. There is some evidence for interviewer experience driving a downward trend in responses, but the pattern is not particularly strong.

Second, even accounting for trends in responses due to interviewer experience, we see a sharp jump in responses right at the time of the intervention. The polynomial fits differ by more than 10 percentage points, and the bin-average confidence intervals barely overlap. The numerical results (column 1 of Table 3.2, panel A) show that this jump is statistically significant: the RD estimate of the treatment effect is 15 percentage points, with a p value of .03.

Another way of assessing the role of interviewer experience is to compare the results with another complex survey module. The questions about sexual activity in the past week

Figure 3.2: Regression discontinuity plots



Notes: The sample is 1,292 sexually active adults who were successfully interviewed at both baseline and endline. HIV transmission risk belief is the perceived probability of contracting HIV from a single unprotected sex act with an infected partner. Sex acts in past week are measured using a seven-day retrospective sex diary.

Table 3.2: Regression discontinuity estimates

	HIV mission Belief	Trans- Risk	Sex Acts in Past Week (from Sex "Diary")	Respondent is Male	Age
A: Discontinuity at Training Session					
RD Estimate	-0.148**		0.476	0.041	-0.876
Standard Error	(0.069)		(0.362)	(0.085)	(1.926)
Conventional p-value	[0.033]		[0.189]	[0.632]	[0.649]
Robust p-value	0.034		0.155	0.698	0.400
Sample:	1,289		1,292	1,292	1,292
Control-group Baseline	X		X	X	X
Treatment-group Baseline	X		X	X	X
Control-group Endline					
Treatment-group Endline					
Observations	1,289		1,292	1,292	1,292
B: Discontinuity at Training Session, Control Group Only					
RD Estimate	-0.019		0.318	-0.004	-1.379
Standard Error	(0.144)		(0.346)	(0.170)	(2.227)
Conventional p-value	[0.896]		[0.358]	[0.981]	[0.536]
Robust p-value	0.776		0.173	0.748	0.386
Sample:					
Control-group Baseline					
Treatment-group Baseline	X		X	X	X
Control-group Endline	X		X	X	X
Treatment-group Endline					
Observations	1,308		1316	1316	1316

Notes: Sample includes 1,376 sexually-active adults who were successfully interviewed at both baseline and followup. Regression discontinuity estimates generated first-degree polynomials and automatic bandwidth selection following Calonico et al. (2014). Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

were collected using a retrospective sex diary originally developed by (Kerwin et al., 2011). In this module, interviewers walked respondents through each of the previous seven days to record details about each sex act on each day as well as other events on that day. The other events included when they woke up and went to sleep, whether they or their partners were menstruating, and alcohol consumption. These other details were included to capture risk factors and to help respondents remember specifics about their sexual activity, similar to an event-history calendar (Belli et al., 2001) or a relationship-history calendar (Luke et al., 2011). This module was complicated to carry out and required the most attention when teaching interviewers to conduct the survey. If the complexity of the questions assessing HIV risk perceptions led to changing response patterns as interviewers gained experience, we might also expect a similar pattern for the sex diary questions.

Panel b of Figure 3.2 presents an RD plot for the number of sex acts in the past week, as reported on the sex diary module. The left side of the graph (before the HIV information training session) shows no clear trend, although a dip is visible just before the training session. Notably, this dip is matched on the right side of the graph, so that the confidence intervals for the bins just before and just after the discontinuity largely overlap. In column 2 of Table 3.2, panel A, we show numeric estimates of the size of this RD. Consistent with the graph, there is a small positive but statistically insignificant difference.

The HIV information training session occurred during a six-day gap in the data collection schedule. To assess whether a break in surveying could be creating the differences in the baseline responses, we look for discontinuities in responses between the end of the baseline survey and the beginning of the endline survey; there was a 10-day break in data collection between the two survey waves. Using the control-group data only, panel c of Figure 3.2 plots an RD for the recorded HIV risk perceptions in the baseline surveys (left side of the graph) versus the endline surveys (right side). The confidence intervals overlap, and the estimated difference (column 1 of Table 3.2, panel B) is nearly zero and statistically insignificant. We see similar null results for the number of sex acts in the past week from the sex diary, suggesting that a break in surveying *per se* does not appear to have effects on the recorded survey responses.

A potential threat to the identification of these RD estimates is that there could have been systematic sorting of respondents around the breaks in data collection. If different kinds of respondents were interviewed just before the training session versus just after, it would be incorrect to attribute the 15 percentage point drop in recorded risk beliefs to the effect of the training session. To test for this sort of systematic sorting, columns 3 and 4 of Table 3.2 present RD estimates for fixed respondent characteristics: gender and age. There are no statistically significant differences in either characteristic for the HIV information training session or for the end of the baseline.

Imbalance

A second potential explanation for the differences between the treatment and control-group beliefs at baseline is imbalance. Although our randomized experiment ensures that the two groups were balanced in expectation, for any given realization of the random assignment it is possible for them to have differences (Frison and Pocock, 1992). Those differences could in turn lead to different beliefs. A particular concern is balance on sexual activity, which is correlated with risk beliefs (Smith and Watkins, 2005). While the sexual activity variables in Panel A of Table 1 are balanced overall, the first five rows all show higher values for the control group than the treatment group. To test for an aggregate balance problem in these variables, we construct an alternate sexual activity index that uses those first five variables alone. The difference is not statistically significant ($p = 0.149$). However, even a statistically-insignificant difference in this variable could lead to substantively-important differences in the belief variables. To mitigate this concern, we control for this alternate sexual activity index in all our regression analyses. Our results do not depend on this choice: the main effects on beliefs from Table 3.1 are barely changed if we drop this control or if we drop the interviewer fixed effects as well.

Another potential source of imbalance concerns variations in religion, ethnicity, and languages spoken across the two groups. HIV risk perceptions and sexual behavior vary widely by religious denomination in Malawi (Trinitapoli, 2009; Trinitapoli and Regnerus, 2006) and ethnicity-specific cultural activities, such as initiation rites, are ways that people learn about sexual health (Munthali and Zulu, 2007). Administering surveys in an unfamiliar language can lead to item nonresponse and systematic measurement error (Andreenkova, 2018). This could be an issue given that all our surveys were administered in Chichewa, but this concern is substantially mitigated by the fact that virtually all our subjects are fluent speakers of either Chichewa or the mutually intelligible language Chinyanja. In the 1998 Malawi census, 96% of households in the study area (TA Mwambo) reported that their most commonly used language was Chichewa or Chinyanja (Minnesota Population Center, 2019).

Table A.5 in the appendix shows balance statistics for specific religious denominations as well as ethnic groups. Panel A shows that although the treatment is balanced in terms of the share of Christians and Muslims (Table A.2 in the appendix), there are important differences across study arms for some specific denominations. In contrast, the treatment is fairly balanced by ethnic group (panel B). However, this pattern may mask potential differences in language abilities within ethnic groups. Because the survey did not ask respondents whether they speak Chichewa at home or whether it was their first language, we use Chichewa-language literacy as a proxy for ability to speak the language. Table A.6 in

the appendix shows balance statistics for literacy in Chichewa by ethnic group. There are no large differences, but the 2 percentage point difference for the “other” group is statistically significant.

To account for potential differences in responses driven by these variations in religion and ethnicity, we add indicators for membership in each group to our regression. We also add indicators for a person being from a given ethnic group and also literate in Chichewa. The results are in Table A.7 in the appendix. The effects on measured risk beliefs are essentially unchanged: they remain statistically significant and are slightly larger, on average.

Spillovers

The similarity in responses between the treatment baseline and control endline surveys implies that interviewer knowledge drives our results rather than some other change that occurred at the time of the training session. This similarity could also have arisen through spillovers: if treatment-group respondents told control-group respondents about the information they learned, then we would expect a fall in control-group beliefs. To test for this possibility, we use social network data to count the number of total friends each respondent has as well as the number they have who live in treatment-group villages. We then estimate the following equation:

$$Y_i = \alpha + \beta T_i + \eta TotalFriends_i + \gamma TreatedFriends_i + \epsilon_i, \quad (3.3)$$

where Y_i is the respondent’s endline risk belief; we also run versions of the regression that break out the spillovers by study arm. Eq. (3.3) identifies spillover effects on endline beliefs because a respondent’s number of treated friends is randomly assigned conditional on their total number of friends (Kremer and Miguel, 2007). We see no evidence of spillovers onto the control group. Spillovers could also have occurred if control-group respondents sought out information about HIV because they were asked about it. We cannot rule out this possibility, but it is unlikely to have generated the observed empirical pattern. To do so, the control group’s information-seeking would need to have led to endline beliefs that were nearly identical to the (measured) baseline beliefs for the treatment group. This is unlikely because the treatment group did not have any time to seek out information about HIV risks prior to answering the baseline questions about risk beliefs.

3.6 Mechanisms

Our results show that being surveyed by a more-knowledgeable interviewer causes a decrease in recorded risk beliefs, and that this effect occurs not just for the risks that the interviewer was directly taught about but also for related risks. We explore three possible

mechanisms for these spillovers between interviewer knowledge and the (recorded) beliefs of survey respondents.

3.6.1 Priming

Given that the surveys involve a face-to-face conversation between respondents and interviewers, interviewer knowledge could affect recorded responses via priming. We find evidence that trained interviewers primed respondents to give answers that matched the exact numbers used in the training. Table 3.3 shows regressions of indicator variables that take a value of 1 when respondents answered exactly 10% for each one of the risk belief questions; this is the exact figure that the interviewer training provided as the true value of the HIV transmission risk per year for unprotected sex. Interviewer training makes respondents more likely to answer exactly 10% for the per-act risk of unprotected risk (column 1), even though that is not the true risk.⁷ We therefore interpret this coefficient as the result of interviewers priming or nudging respondents toward lower responses to all risk belief questions, not just the one corresponding to the information treatment. However, we do not see an increase in reporting an answer of exactly 10% in column 2 (the annual risk), where it is the correct answer. A potential explanation is that respondents have extremely high priors for this figure: the average risk belief is 93%. In columns 3 and 4 (condom-protected sex risks), we see slight reductions in the chance that people report exactly 10%. Because those questions immediately followed the questions about unprotected sex risk, this could be explained by respondents updating their risk beliefs consistently: if condoms lower the risk by a factor X , and the unprotected-sex risk is 0.1, then the condom-protected risk is $0.1X$.

These results are consistent with the literature on priming and anchoring, which has shown that mentioning numbers will induce people to give answers to subsequent questions that are more similar to those numbers (Newell and Shanks, 2014). This can happen by directly suggesting a potential answer, exposing respondents to peers' responses (Tversky and Kahneman, 1974), or even mentioning totally unrelated numbers (Chapman and Johnson, 2002; Mussweiler et al., 2000).⁸ Although all three priming pathways are possible in our context, the first is the most likely. Interviewers were trained to encourage respondents to answer even if they were not sure, and one way they might have done so is to ask leading questions like, "Do you think it might be $X\%$?" It is likely that interviewers who were exposed to the training were more likely to suggest 10% as a possible answer.

⁷The information treatment mentioned only the annual transmission risk for unprotected sex, and the figure provided for the true risk was 10%. The true transmission risk per act is approximately 0.1%.

⁸Similarly, (Delavande et al., 2017) found that subjective beliefs are affected by the exact wording of the question—in that research, framing a question as being about mortality versus survival.

Table 3.3: Treatment Effects on Answering Exactly 10%

	(1)	(2)	(3)	(4)
	Outcome: Recorded belief is exactly 10%			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment	0.0427*** (0.0117)	0.00665 (0.00601)	-0.0507*** (0.0189)	-0.0201 (0.0171)
Control-group mean	0.022	0.005	0.100	0.100
Control-group SD	0.146	0.068	0.300	0.300
Observations	1,282	1,277	1,284	1,277

Notes: All regressions control for stratification cell and interviewer fixed effects, as well as the alternate sexual behavior index from Appendix Table A.2. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

3.6.2 Encouraging Guesses

Another opportunity for interviewer knowledge to affect respondents' recorded beliefs comes from the structure of our subjective belief elicitation questions. These were designed such that whenever respondents answered 50% to any risk belief question, a follow-up question asked whether they really thought the answer was 50%, or whether they were just unsure. If respondents said they were just unsure, they were asked for their best guess. This approach was adapted from the U.S. Health and Retirement Survey (HRS), with the goal of reducing the use of 50% as a proxy for respondent uncertainty; see Hudomiet et al. (2011) for a discussion of this technique.

These follow-ups initiated another interaction between the interviewer and respondent, creating an additional opportunity for interviewer knowledge to spill over onto survey responses. Table 3.4 shows our exploration of that additional interaction, in the case of per-act transmission risks for unprotected sex. We create indicator variables for when respondents answered 50% (column 1) and for when they changed, decreased, or increased their answer when asked the follow-up question (columns 2–4). Columns 5 and 6 also show whether answers decreased or increased, restricting the sample to those respondents who initially answered 50%.

Respondents in the treatment and control groups are equally likely to answer 50% and

Table 3.4: Priming by interviewers when initial answer was 50%

	Answer = 50%	Changed Answer	Decreased Answer	Increased Answer	Decreased Answer	Increased Answer
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0161 (0.0223)	-0.0126 (0.0122)	0.000651 (0.00704)	-0.0133 (0.00872)	-0.00832 (0.0472)	-0.115* (0.0630)
Conditional on Initially Answering 50%	N	N	N	N	Y	Y
Control-group mean	0.115	0.042	0.008	0.034	0.0676	0.297
Control-group SD	0.319	0.201	0.088	0.182	0.253	0.460
Observations	1,285	1,285	1,285	1,285	159	159

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and end-line. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

equally likely to change their answer after the follow-up (column 2). However, respondents who were exposed to this additional interaction were significantly less likely to increase the answer after the follow-up question when they were interviewed by a trained interviewer, as shown in columns 4 and 6 of Table 3.4. Column 6 shows that conditional on initially answering 50%, respondents exposed to informed interviewers were almost 20 percentage points less likely to increase their answer. This magnitude is large, considering that only about 30% of those in the control group increased their answers after the follow-up question.

We interpret these results as additional evidence that interviewers who were exposed to the information treatment influenced the responses given by communicating the follow-up question in a way that nudged or primed respondents not to increase their answers. Such subtle communication could be anything from a change in tone of voice or body language to the choice of words.⁹ A specific possibility is that instead of asking whether the number could be more or less than 50%, they asked only if it could be less. We do not believe interviewers did this intentionally: they knew that the purpose of the intervention was to study the respondents' knowledge and behaviors, and their incentive was to record information accurately. Rather, we believe that interviewers inadvertently nudged respondents

⁹Ideally, we would have directly observed some interviews to measure the microprocesses that drove the knowledge spillovers. We do not do this for two reasons. First, the study was not designed to measure these spillovers. Second, direct participation in the survey by outsiders, especially White foreigners, can itself affect respondents' behaviors (Cilliers et al., 2015).

toward lower answers. Equally possible is that interviewers who were not exposed to the information treatment nudged respondents to provide higher answers, potentially stemming from their own (high) beliefs prior to the information treatment.

3.6.3 Interviewer Knowledge and Respondent Priors

If interviewer knowledge spillovers indeed operate through nudges and priming that take place during the survey interview, we would expect the effects to be smaller for respondents who are more confident in their beliefs. Because our outcomes are measured at the baseline survey, we do not have direct measures of respondents' beliefs in the absence of the knowledge spillovers. However, some of their other characteristics may be useful proxies. Table 3.5 examines treatment effect heterogeneity by a range of respondent characteristics, estimated using Eq. (3.2). We observe significant heterogeneity by years of schooling and total assets.

These characteristics are correlated with one another, and thus we may be finding spurious heterogeneity by some characteristics due to omitted variable bias. Therefore, we include all eight interactions in column 9 and add interactions with additional characteristics as well in column 10. In our preferred specification, column 10, only the interaction between treatment and years of schooling remains significant.¹⁰ The positive coefficient for years of schooling means that the main treatment effect (the effect of having a more knowledgeable interviewer) is smaller in magnitude for those with more education.

To further explore this finding, we run another set of regressions with the dependent variable being beliefs about HIV transmission risks per unprotected sex act and the independent variables including the treatment interacted with seven different measures of schooling: years of schooling, having completed at least Form 1 or Form 2, and having completed Forms 1–4.¹¹ The results suggest that the treatment effects are lowest for people who reached Form 2. Although our survey respondents were all adults, most had not attended secondary school. Just 20% had completed Form 1, and only 17% had completed Form 2.

Form 2 is the point at which students in Malawi are most exposed to information on HIV transmission and sexual health.¹² NGOs in Malawi also tend to target students of

¹⁰The standard errors in this specification may be overstated because of multicollinearity between age, years of schooling, and years sexually active. The condition number for the three variables is nearly 18; a figure greater than 10 can indicate that coefficients are unstable. However, the variance inflation factors are all well below the usual cutoff of 10.

¹¹Form 1 in Malawi is the equivalent of ninth grade in the United States.

¹²HIV education was moved from other subjects into a course called life skills in the early 2000s (Chamba, 2009). When this change was initially rolled out in 2001–2002, HIV was included only in the secondary school life-skills curriculum. The current life skills curriculum in upper primary school (grades P5–P8) is supposed to include HIV education, but there are many constraints to implementation (Chirwa and Naidoo, 2014).

Table 3.5: Heterogeneity in treatment effects

Outcome: HIV transmission risk belief (per-act, unprotected)										
Treatment	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	-0.0753*** (0.0191)	-0.0745*** (0.0191)	-0.0756*** (0.0191)	-0.0769*** (0.0194)	-0.0770*** (0.0200)	-0.0737*** (0.0191)	-0.0746*** (0.0191)	-0.0804*** (0.0194)	-0.0891*** (0.0210)	-0.101*** (0.0259)
T × (Age)	0.000458 (0.00177)								-0.00230 (0.00469)	-0.000716 (0.00517)
T × (Male)		-0.0115 (0.0307)							-0.0558 (0.0338)	-0.0681* (0.0372)
T × (Years of Schooling)			0.0112*** (0.00397)						0.00768 (0.00501)	0.0102* (0.00543)
T × (Years Sexually Active)				0.000258 (0.00175)					0.00422 (0.00482)	0.00290 (0.00506)
T × (30 Day Income)					0.0162 (0.0125)				0.00771 (0.0133)	0.00793 (0.0135)
T × (Total Assets)						0.0144** (0.00700)			0.00694 (0.00887)	0.00704 (0.00884)
T × (Any Sex in Past 7 Days)							0.0160 (0.0353)		0.00995 (0.0367)	-0.0255 (0.0462)
T × (Numeracy Score)								0.0223 (0.0165)	0.0130 (0.0191)	0.0175 (0.0190)
Interactions w/other covariates†	N	N	N	N	N	N	N	N	N	Y
Control-group mean	0.827	0.827	0.827	0.827	0.827	0.828	0.827	0.827	0.827	0.832
Control-group SD	0.264	0.264	0.264	0.264	0.263	0.264	0.264	0.264	0.264	0.260
Observations	1,282	1,282	1,282	1,267	1,197	1,281	1,282	1,282	1,182	1,166

Notes: All regressions include controls for stratification cells, interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline; 120 of these have missing data for at least one of the covariates. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

† Other covariates include ravens test score, immediate word recall, any sex in the past 30 days, total sex acts in the past 7 days, an indicator for respondent changing their answers, and categorical indicators for ethnic group.

this age for HIV-prevention interventions; in other African countries, it is also common to target HIV-prevention campaigns toward students early in secondary school (Gallant and Maticka-Tyndale, 2004). The narrative to which students in Malawi are exposed in these lectures and courses is that HIV is highly contagious, which should lead to high risk beliefs and high certainty about those beliefs.¹³

Other people have likely heard about the information provided to students in Form 2—for example, by hearing about it from their friends. Other institutions such as NGOs and church groups attempt to teach about HIV transmission, and so other people may also have strong priors about HIV transmission risks. However, direct exposure to this information in school could lead people to be more certain about it and thus less susceptible to nudges by interviewers.¹⁴

3.7 Preventing and Mitigating Interviewer Knowledge Effects

Researchers can combat the potential spillover of interviewer knowledge onto recorded subjective beliefs in a variety of ways. First, researchers can try to alter how interviewers are recruited and trained. Recruitment matters because these spillovers can occur whenever interviewers and respondents differ in their knowledge levels. Researchers should recruit interviewers who match the respondent population as closely as possible, particularly in terms of education and exposure to information relevant to the survey questions. This will prevent knowledge effects, assuming that the effect we measure is driven by gaps between interviewers' and respondents' knowledge. Controlling for interviewer fixed effects can help eliminate the effects of any remaining knowledge differences by purging the results of any interviewer-specific patterns. The survey design should also include exact scripts for asking belief questions to minimize selective nudges by the interviewer. Training sessions should emphasize the potential for these spillovers and coach interviewers on how to avoid them.

Second, the problem can be tackled through changes in the design of experiments when

Based on conversations with Ministry of Education officials in Malawi in 2012, at that time HIV education was done only in secondary schools. An examination of the textbooks for the secondary-school life skills courses revealed HIV content in all four grades, but HIV transmission risks were covered only in Form 2 (Kadyoma et al., 2012).

¹³This idea is supported by the correlation between risk beliefs and schooling for the control group: more schooling is associated with higher priors.

¹⁴Another potential explanation for the fact that more-educated people respond less to the treatment is that education may make people more confident and better able to withstand influence from outsiders. We cannot directly test this explanation against the effect of HIV education on the strength of people's priors.

studies involve information treatments. Possible solutions include either running the baseline surveys simultaneously across the treatment and control groups or separating the information treatment from the survey data collection entirely. Each strategy has important potential drawbacks. Running simultaneous surveys across study arms creates the possibility that respondents will be given the wrong version of the survey and thus be unintentionally exposed to the information treatment, creating a far worse contamination problem. Running the information treatment separately from the survey—for example, via village meetings—can make it difficult to prevent nontargeted people from receiving the information. For example, diagrams that are distributed as handouts could potentially make their way into the hands of control-group subjects. Thus incorporating information treatments into surveys is likely to *minimize* information spillovers, not exacerbate them, but at the cost of yielding potentially biased measurements of respondent beliefs. If accurate measures of respondent beliefs are not a main goal of the study, this may be an acceptable risk. For example, if the goal of an experiment is to see how much an information treatment shifts behaviors, then mismeasured beliefs are not a problem even if they affect only one of the study arms. Even when examining treatment effects on risk beliefs is an important goal of the study, interviewer knowledge contamination is a problem only if it interacts with the actual treatment. Apart from information experiments, knowledge spillovers are likely to occur simply because interviewers differ in their knowledge and beliefs. Providing interviewers with a basic level of knowledge about important survey topics could help them do their jobs better and lead to better-quality data.

A third solution to this issue is to collect subjective expectations in a way that avoids any direct interaction with interviewers, such as by using computer-assisted self-interviewing (CASI). This would eliminate any possibility of interviewer knowledge spilling over onto respondents' survey responses. Research on CASI has shown it to be effective in low-literacy settings (Hahn et al., 2003; van de Wijgert et al., 2000). Important limitations exist, however: participants may not be able to clarify questions (Group, 2007), computers may be received with suspicion in certain settings (Hewett et al., 2004; Mensch et al., 2003), and bystander presence might affect results and should be recorded or controlled (Aquilino et al., 2000 May-Jun). Potdar and Koenig (2005) argued that CASI will not yield more-honest answers if people are not comfortable using computers. In low-literacy settings, audio computer-assisted self-interviewing (ACASI) may work well; Rumakom et al. (2005) showed that it outperforms self-administered questionnaires. However, Soler-Hampejsek et al. (2013) tested ACASI for collecting data on sexual activity in Malawi and find that it still leads to high rates of inconsistency in responses. Similarly, Mensch et al. (2008) showed that face-to-face interviews generate lower rates of inconsistencies in responses than ACASI, as well as stronger correlations between reported sexual behavior and biomarkers for HIV

infection. To improve the quality of subjective expectation data in developing countries, more work on adapting CASI and ACASI to overcome these limitations is needed. One promising approach is to use computer tablets to have respondents play simple games that convey information; Tjernström et al. (2019) showed that this approach is successful in a population of Kenyan farmers.

3.8 Conclusion

Leveraging a randomized experiment that used interviewers to measure subjective HIV transmission risks and provide information to treatment group participants, we find that interviewer knowledge affects the recorded values of survey respondents' subjective beliefs. This information spillover occurs not only for the information directly given to the interviewers but also for other related risks.

We identify several channels through which these effects happen. They are evident at various points in the survey, including the follow-up questions triggered by respondents answering 50%. This result suggests that additional interactions between interviewers and respondents present the potential for more spillovers. Our evidence suggests that interviewer effects work via priming or nudging rather than interviewers directly revealing information.

We find that interviewer effects are weaker for more-educated people, possibly because those respondents received information about HIV transmission directly at school. This could make them more certain about their prior beliefs than those who heard information secondhand, even if the level of those beliefs is not different across education levels.

Our results have important implications for demographers as well as other social scientists who study subjective expectations themselves or phenomena that are driven by people's subjective beliefs. Subjective expectations have proven to be useful tools for understanding and forecasting the main demographic processes of fertility, mortality, and migration, but these uses rely on being able to measure them correctly. Researchers need to be aware of the possibility that interviewer knowledge will spill over onto respondents' recorded beliefs, which could have substantive effects on results that use those beliefs. Although our findings relate to HIV risk beliefs, interviewer knowledge effects could occur for any subjective expectation where the interviewer knows more than the respondent—including other diseases such as Ebola or COVID-19, and also other domains where subjective expectations play a role, such as conception probabilities, mortality rates, and the returns to migration. Our results were obtained in a setting where interviewers were evaluated based on recording responses correctly; these results may not generalize to settings where interviewers have some interest in recording specific responses.

We suggest methods for correcting this problem at several points during the course of a

research project. These include mindful recruiting of interviewers to match the knowledge levels of the respondent population, emphasizing the potential for spillovers in training, and designing the experiment in such a way that interviewers survey both study arms while having the same information set. The most promising way to avoid interviewer knowledge effects is to collect data via CASI or ACASI to reduce the scope of interaction between respondent and interviewer, but both methods have issues with data quality. Interviewer knowledge effects are therefore likely to remain an issue for measuring subjective beliefs in developing-country settings for the foreseeable future.

Chapter 4

Soup Kitchens and Food Security: The Case of Mexico's Crusade Against Hunger

4.1 Introduction

Food security and nutrition are two important aspects of development. At the World Food Summit of 1996, the following definition of food security was adopted: “Food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life.” Famines, hunger and malnutrition have been a major concern for high and low income countries, and increasingly so as we expect climate change to worsen food access for many people. Researchers have found various negative effects of food insecurity -short and long term pervasive effects on variables ranging from low productivity to decreased child health (Gundersen and Kreider, 2008), and reduced non-cognitive development (Howard, 2011).

This has led governments all around the world, as well as international organizations, to spend large amounts of resources in programs aimed at ensuring food security for their population. Examples of such programs include soup kitchens and food banks¹. These programs, such as TEFAP by the USDA, have been promoted as means to alleviate food insecurity. However, little evidence has been presented about the effect of these interventions

¹For the purpose of this chapter, food banks and soup kitchens are defined as spaces where food is provided either for free or at a subsidized price with the purpose of aiding people who face economic hardship and other issues that hinder their food access. Soup kitchens serve prepared food on the premises while food banks hand out food products to be prepared and consumed elsewhere.

in individual or aggregate outcomes.

In the absence of randomized control trials, it has been difficult to evaluate the effectiveness of soup kitchen and other in-kind food provision programs on food security, given the endogeneity of treatment as well as participation misreporting. There is some evidence that SNAP and School lunches improve food security in the US (Kreider et al. (2012), Gundersen et al. (2012), Kreider et al. (2016)). There is also evidence from a small-scale program (Mousa and Freeland-Graves, 2019) that shows positive effect of in-kind food provision on food security. Scholars have also written about how emergency vs. sustained provision programs may have very different implications, since food insecurity is a complex phenomenon, thus suggesting chronic and acute food insecurity should be tackled differently.

While subsidized food supply is attractive as a potential solution to food insecurity, there are conditions under which soup kitchens would not necessarily improve food security or nutrition. A few examples of such conditions include poor targeting, implementation issues, crowding out, or low quality food provision. Poor targeting could mean that the people who actually attend the soup kitchen are not the food-insecure. Crowding out refers to the program acting as a substitute of previously existing ways of accounting for food insecurity such as previously existing soup kitchens, social cooperation, or other transfers or donations. In either of these cases, food security would not improve with a soup kitchen the program. Additionally, even if the target population is the treated population and they were not crowding out other previous mechanisms dealing with food insecurity, if the soup kitchens do not provide foods for an adequate nutrition, they would not help eliminate many of the effects of food insecurity and malnutrition. In other words, the fact that some people are attending soup kitchens is not sufficient for overall food security to improve, given the current and widely accepted definition of food security.

In this chapter, I estimate the effects of having access to a soup kitchen on food security, using a program implemented in Mexico starting on 2013. I observe a sample of Mexican households from 2012-2014, which allows me to evaluate only short term effects of the program. I take a difference-in-differences approach to estimate the effect of the program on six different measures of food security, standard in the literature. I compare municipalities with a program-funded soup kitchen to ex-ante similar municipalities without one. Using the Nationally representative Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH), which is a repeated cross-section household survey, I am able to control for proxies of social capital within communities. I also link electoral data at the municipal level to explore differences across municipal and state political alignment to the Federal government.

I find that the presence of a soup kitchen in the municipality shows no mean effect on any one of my measures of food security. To explore the possibility of program effects being concentrated on the left tail of the food security distribution, I perform the same analysis

for subsets of the most food insecure individuals within municipalities. I find some evidence that there is indeed a larger effect on the most food insecure population, however, my study is under-powered for this analysis.

Exploiting the rules of the program, I perform robustness checks using municipal political alignment with the federal government as a triple difference. My results suggest that it is not through perverse political alignment incentives that the program has not shown mean effects.

My contribution is twofold: evaluating the effect of the soup kitchen program on the mean and on lower percentiles of the food security distribution, and exploring possible channels through which the program has failed to show positive mean effects on food security. The conclusions of this study are important pieces of information for policy makers. Moreover, the caveats of this study show that there is a pressing need for better data of the target population, including take-up, in order to thoroughly evaluate these programs. Finally, this study shows that sometimes policymakers may target objectives that are not the most pressing to beneficiaries.

The remainder of the chapter is organized as follows: in Section 4.2 I give an account of the structure of the program and the context of its implementation. In Section 4.3 I present and describe the data. Section 4.4 contains the results for my initial difference-in-differences estimation and further analyses investigating effects on different parts of the distribution of food security, effects on food expenditure, and the political-incentives channel. Lastly, I conclude on Section 4.5.

4.2 Background

The Mexican government announced in 2012 that they were launching a set of programs they called their “Crusade Against Hunger”, implemented by the social development agency of the federal Mexican government (SEDESOL). The Crusade consisted of more than 60 programs, many of which already existed (such as Prospera, formerly known as Oportunidades), but would receive new funding under the Crusade’s budget. The goals of the Crusade were varied, but included “zero hunger”, or adequate nutrition for those in extreme multidimensional poverty and without adequate food access, as defined by CONEVAL, the decentralized agency in charge of poverty measurement and policy evaluation in Mexico. Other goals of the Crusade were to eliminate acute child malnutrition, increase income for small agricultural households and decrease post-harvest food waste.

The soup kitchen program I study in this chapter began with the Crusade. Its announcement was accompanied by the rules under which the program would operate (SEDESOL, 2013). Once a soup kitchen was set up, it would feed up to 120 people, twice a day, five

days a week. Those who receive meals at the soup kitchen had to be part of the “at risk population”, meaning they should be in a state of poverty, malnutrition or being at risk of experiencing it. The Crusade stated that children, pregnant women and the elderly should be prioritized, however there were no formal mechanisms to enforce these priorities.

Soup kitchens are staffed by volunteers from the community, and some members who may offer their labor in exchange for meals. One day working at the kitchen can be exchanged for meals for up to three people that day. For the operation of the soup kitchen, the federal government provided up to around \$7,000 USD in kitchenware² and sends a monthly bundle of non-perishables valued up to \$4,000 USD. The federal government also provided training for the kitchen volunteers through the army, navy and other government agencies. Additionally, kitchens could charge up to \$1 USD per day per person, to account for either buying perishable goods or paying electricity or water bills. An exploratory study (CONEVAL, 2015) showed that soup kitchens generally didn’t charge or charged well under the established limit.

Setting up a soup kitchen, however, required several steps. First, Crusade municipalities were selected, on the basis of their poverty indicators and other features³. Only these municipalities were eligible for accessing Crusade funds. Afterwards, in communities within the selected municipalities, a committee must then be democratically⁴ elected. The committee must apply for funding to set up a soup kitchen. In this application, the committee must provide a list of people who would benefit from the soup kitchen, who are part of the target population and live within the community. The application must also include the list of volunteers who will staff the kitchen, and it must specify the physical space in which the soup kitchen will be set up. There is a preference in the program for setting up kitchens in publicly owned spaces (except schools, which are not allowed to house soup kitchens), but could be private if no public spaces are available. In this case, the owner must agree in writing to lend the building to the program for at least a year. The space must have running water and must be big enough to house a dining space for up to 120 people and the kitchen space, among other restrictions (SEDESOL, 2013).

There are some potential issues with this program design. One of them is not targeting the poorest communities, where food security may be more pressing. It is conceivable that poorer communities find it harder to set up a committee, find a space to set up the kitchen and request funding, with the bureaucratic process it entails. In fact, the literature suggests that communities with lower social capital are more at-risk of food insecurity (Martin et al.,

²Depending on the type of kitchen: rural or urban

³CONEVAL (2014) noted that out of the first 400 municipalities selected, 16 did not satisfy the strict indicator selection

⁴There are no guidelines within the program description as to what this means exactly, nor are there ways to enforce it

2004). Considerable support from local authorities is also needed, in particular with the provision of the physical space. The design of the program could thus be making it harder for the poorest, most food insecure communities. Regarding targeting, another issue is that the list of beneficiaries is constructed by the community committee and there is no formal way of verifying that those attending are part of the target population.

Another potential problem is that since fresh food is not directly provided by the program, the nutritional value of the meals served is not directly observed or guaranteed. Although training is provided to the volunteers, there is no control over what is served and no direct instruction of how it should be nutritionally composed. Food security improvement would require that the people who are vulnerable actually attend the soup kitchens, and receive nutritious meals.

In the first year, SEDESOL funded 500 soup kitchens in the state of Guerrero and spent around US\$12,558,000. The following year, the expenditure in the program had risen to US\$77,775,000 and the number of soup kitchens exceeded 3,156 (CONEVAL, 2015). By 2015, SEDESOL announced that 3.4 million Mexicans benefited from the programs of the Crusade. They argued that soup kitchens and other Crusade programs implied that more people have secure access to quality food (SEDESOL, 2015).

There has been some journalistic work on quality and variety of the food provided by the soup kitchens, as well as some anecdotal testimonies of beneficiaries (for example, Animal-Politico (2016)). CONEVAL carried out an exploratory study in 2014 showing some interesting results and concluding that the soup kitchen program design was not conducive to long term benefits. SEDESOL regularly presents its own back of the envelope estimation of number of people benefited. However, before the present chapter, no empirical evidence had been presented on how the soup kitchen program had in fact impacted food security for impoverished Mexicans.

4.3 Data and Empirical Strategy

To study whether this program had any effects on food security in the municipalities where it was set up, I use two cross-sections of the ENIGH: 2012 and 2014. The ENIGH is a household level survey. It is a very rich data set that has socio-demographic, expenditure and other data for each member of the household. It allows me to control for income, a proxy for social capital, characteristics of the head of the household and others.

There is currently no publicly available data set that lists the operating soup kitchen and when they opened. To identify municipalities that had a soup kitchen open between 2012 and 2014, I take a list of municipalities with a soup kitchen operating during 2014, and a list of 2013 Crusade municipalities, exploiting the fact that only the original Crusade

municipalities were eligible to open a soup kitchen in 2013.

In defining my treatment, I therefore assume that no soup kitchens closed between 2013 and 2014, and that no soup kitchens were opened in non-2013 Crusade municipalities in that year. Municipalities labeled as treated in my analysis are municipalities that had a soup kitchen in 2014, and were also eligible to open a soup kitchen in 2013. This assumption is untestable for now, but it is the best use of the available data.

For my analysis I compare municipalities in which a soup kitchen operated to those in which none operated during this time even though they were eligible. As dependent variables, I use a set of six food security related indicator variables from ENIGH. Table 4.1 shows the text of the questions. I construct binary variables, where a value of 1 means that the household answered ‘yes’ for that question, which indicates the household was at risk of or experienced food insecurity in the last three months.

Each one of the dependent variables is an imperfect measure reflecting different aspects of food security, none of them by themselves fully capture the complex phenomenon. Rather than classifying my sample in food secure vs food insecure by some aggregation of all variables (Such as in Wilde and Nord (2005), for example) and choosing a cutoff point, I perform my analyses for each one of them and present all results, to give a more detailed picture of the effects of the program, following the advice of Maxwell et al. (2014).

All the dependent variables are related to having enough, varied food in a fixed period of time. For this reason, this data cannot distinguish between chronic and temporary food insecurity as the indicator variables turn on with one event occurrence, and we observe them only twice in time. In the present chapter I focus on food security for the whole household, taking the responses of the household head.

Table 4.1: Measures of Food Security

Q1	Do you ever worry about running out of food due to lack of resources.
Q2	Did you ever run out of food due to lack of resources.
Q3	Did you ever not have enough resources to have healthy and varied meals.
Q4	Did you ever eat low variety meals due to running out of resources. †
Q5	Did you ever skip a meal due to lack of resources. †
Q6	Did you ever eat less than you think is necessary due to lack of resources. †

Notes: †These questions were asked specifically about adults in the household.

Table 4.2 presents summary statistics split between households in crusade municipalities with at least one soup kitchen and households in crusade municipalities without any, in the 2012 wave. Standard errors are clustered at the municipality level. Expenditures

and income are calculated in Mexican Pesos, per month. Male headed households, Literacy, and Indigenous show the fraction of households headed by men, the fraction of the population that can read and write, and the fraction that self-identifies as indigenous. Occupied Members indicates the number of household members that perform paid work outside the home. The only significant difference is in the age of the household head. Municipalities with soup kitchens seem to have younger heads of household on average.

Table 4.2: Summary Statistics for Treated and Non-treated Households

	(1) Control	(2) Treatment	(3) Difference
Sum of Food Security Questions	1.720 (2.11)	1.85 (2.13)	0.136
Household Head Education	5.590 (2.72)	5.17 (2.66)	-0.419
Household Head Age	49.08 (15.67)	47.51 (15.67)	-1.579**
Female Headed	0.27 (0.44)	0.25 (0.43)	-0.02
Can Read and Write	0.87 (0.23)	0.85 (0.24)	-0.016
Occupied Members	1.7 (1.08)	1.72 (1.11)	0.015
Per Capita Income	14,763.35 (27,695.51)	12,236.22 (18,676.02)	-2,527.12
Number of Members	3.77 (1.91)	3.82 (2.09)	0.052
In Kind Transfers Received	653.81 (3,303.04)	503.35 (2,354.53)	-150.46
Total Expenditures	34,708.73 (34,751.34)	30,243.85 (30,077.07)	-4,464.89
Food Expenditures	8,361.56 (6,209.5)	8,293.54 (6,565.29)	-68.016
Native Population	0.31 (0.45)	0.35 (0.46)	0.037
Female Members	1.95 (1.24)	1.93 (1.32)	-0.019
N	2,121	1,617	

Notes: Standard errors, clustered by municipality, are shown in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure 4.1 shows the share, by treatment group, of households with different values of the sum of all six food security binary variables for households in crusade municipalities, in the 2012 wave of the ENIGH. A value of zero indicates that the household head answered ‘no’ to all questions in Table 4.1, while a value of 6 means the household head answered ‘yes’ to all of them. We can see that the most common value in both sets of households is zero, which reflects the fact that the phenomenon they describe is not experienced by all households. However, about half of the households in each group answer yes to at least one of the questions, highlighting the pervasiveness of food insecurity in this context. There is no average difference at baseline in the value of the sum for treated and untreated households, as shown in 4.2.

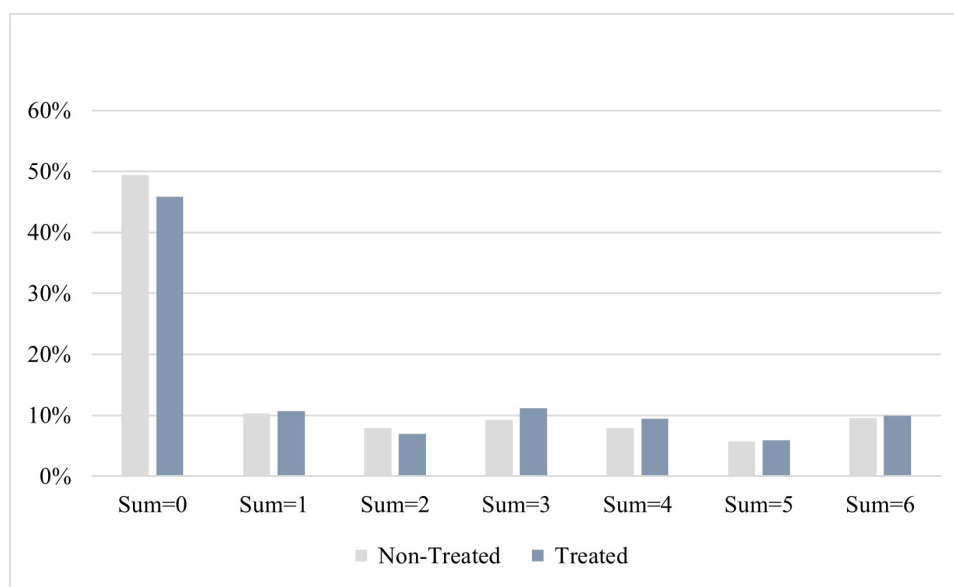


Figure 4.1: Distribution the Sum of all Food Security Questions at Baseline for Treated and Non-treated Households

In all specifications, I also control for a proxy of social capital is constructed as an index of five different answers from the ENIGH questionnaire. The questions are related to how easy the household believes it would be to obtain something in case of hardship. For example, how hard would it be for them to borrow a month’s salary if they needed it or how hard would it be for them to get help to find a job. The answers range from very easy to very hard, or impossible. The higher the value, the easiest each one would be, so the higher the index, the more social capital. This measure will be used in all specifications in the following section as a proxy for the value of social capital. Table 4.3 shows the text of each question.

Table 4.3: Social Capital Proxy Questionnaire

If you needed to, how hard would it be to obtain the following
Borrow your household's monthly income
Someone to care for you if you were sick
Help getting a job
Someone to accompany you to the doctor's office
Funds for community investments
Someone to help you care for the children in this home

I estimate the following equation, where T is either an indicator variable for municipalities in which soup kitchens open, or how many soup kitchens there were. A is an indicator for the year 2014, and DID is the product of the two. X_{it} is a vector of household level controls, and Y_{itm} is either a dummy variable for answering 'yes' to the corresponding question about food security, or the sum of dummies for all questions.

$$Y_{itm} = \gamma T_m + \eta A_t + \beta DID_{mt} + \delta X_{itm} + \epsilon_{itm} \quad (4.1)$$

4.4 Results

Table 4.4 shows the results of estimating Equation 1 with a binary treatment definition, on the sample of all eligible municipalities. We can see that the difference in difference coefficient is not significant for any of the dependent variables. Furthermore, some of the questions show positive coefficients (Q1, Q2, Q5), while some show negative ones (Q3, Q4, Q6). for none of the specifications.

It is possible that the effects of this program are concentrated on the lower portion of the food security distribution, and the effects are too diluted if we estimate them in the entire population. After all, this is a program designed for the most food insecure, and about 50% of the sample answers no to every question regarding food security. For this reason, I estimate the same equation but this time in a restricted sample that includes only households in the most ex-ante food insecure municipalities. Because the ENIGH is not a panel, I cannot create a sample of the most ex-ante food insecure households. Therefore, what I do is focus on households within the most ex-ante food insecure municipalities.

To select the most food insecure municipalities, I construct a food insecurity index, using the answers of the 6 questions in Table 4.1. I then select municipalities that were in the highest 10% of the food insecurity index distribution, in other words, the most ex-ante food insecure municipalities out of the eligible municipalities. Table 4.5 shows the results for

Table 4.4: Estimation of Equation 1 with Binary Treatment on Measures of Food Security

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q6	(7) Sum
DID	-0.0002 (0.0176)	-0.0150 (0.0112)	0.0005 (0.0116)	0.0139 (0.0120)	-0.0044 (0.0111)	0.0072 (0.0094)	0.0020 (0.0043)
After	-0.0029 (0.0098)	0.0037 (0.0076)	0.0023 (0.0087)	-0.0103 (0.0070)	0.0061 (0.0074)	-0.0007 (0.0060)	-0.0017 (0.0027)
T=SK Open	-0.0085 (0.0164)	0.0129 (0.0116)	-0.0008 (0.0141)	0.0014 (0.0124)	-0.0094 (0.0138)	0.0028 (0.0091)	-0.0016 (0.0043)
Control group mean	0.4493	0.1438	0.3470	0.3541	0.1565	0.2645	1.7152
Control group standard deviation	0.4975	0.3510	0.4761	0.4783	0.3634	0.4412	2.1106
Observations	13,246	13,246	13,246	13,246	13,246	13,246	13,246

Notes: Controls include state fixed effects, an index of social capital, percapita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table 4.5: Estimation of Equation 1 with Binary Treatment on Measures of Food Security for the Restricted Sample

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q6	(7) Sum
DID	0.0375 (0.0711)	0.0074 (0.0749)	-0.0327 (0.0411)	-0.0219 (0.0497)	0.0216 (0.0765)	-0.0105 (0.0489)	0.0014 (0.0194)
After	0.0097 (0.0287)	-0.0228 (0.0602)	-0.0031 (0.0304)	0.0040 (0.0431)	0.0109 (0.0604)	0.0035 (0.0212)	0.0023 (0.0133)
T=SK Open	-0.0122 (0.0321)	-0.0172 (0.0572)	0.0679 (0.0429)	0.0425 (0.0348)	-0.0186 (0.0731)	-0.0549* (0.0257)	0.0075 (0.0125)
Control group mean	0.7638	0.5039	0.6693	0.6535	0.4567	0.6142	3.6614
Control group standard deviation	0.4264	0.5020	0.4723	0.4777	0.5001	0.4887	2.3611
Observations	666	666	666	666	666	666	666

Notes: Controls include state fixed effects, an index of social capital, percapita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.6: Estimation of Equation 1 with Continuous Treatment on Measures of Food Security

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q6	(7) Sum
DID	-0.002 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000* (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.000)
After	0.004 (0.009)	0.002 (0.007)	0.001 (0.007)	-0.009 (0.007)	0.002 (0.006)	0.000 (0.005)	0.000 (0.002)
T=Number of SK	0.001 (0.001)	0.002** (0.001)	-0.001 (0.001)	-0.002*** (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.000)
Control group mean	0.449	0.144	0.347	0.354	0.157	0.264	1.715
Control group standard deviation	0.498	0.351	0.476	0.478	0.363	0.441	2.111
Observations	13,246	13,246	13,246	13,246	13,246	13,246	13,246

Notes: Controls include state fixed effects, an index of social capital, percapita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

the estimation on this sample. Even for this restricted sample, I estimate no statistically significant effects for any of the measures of food security, and the sign of coefficients is still varied.

Using the data I have on how many soup kitchens were set up in each municipality, I estimate Equation 1 again, this time with a continuous treatment: the number of soup kitchens that were set up in that municipality. Tables 4.6 and 4.7 show estimations on the entire sample and the sample of most food insecure respectively. The tables show that only for the most food insecure, an additional soup kitchen would decrease the sum of all food security questions (Column 7), thus showing an improvement of food security in that section of the population. The magnitude is not large, and the estimated coefficient is only significant for the sum, and Question 3 (Column 3) which refers to having varied and healthy meals.

Table 4.7: Estimation of Equation 1 with Continuous Treatment on Measures of Food Security for the Restricted Sample

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q6	(7) Sum
DID	-0.0109 (0.0077)	0.0059 (0.0079)	-0.0116* (0.0057)	-0.0022 (0.0050)	0.0058 (0.0066)	0.0093 (0.0053)	-0.0038* (0.0019)
After	0.0574 (0.0446)	-0.0232 (0.0527)	0.0120 (0.0295)	0.0061 (0.0377)	-0.0150 (0.0532)	-0.0228 (0.0343)	0.0145 (0.0137)
T=Number of SK	-0.00413** (0.0018)	0.00739*** (0.0009)	-0.0037 (0.0021)	0.00258*** (0.0007)	-0.00791*** (0.0021)	0.00509*** (0.0012)	-0.0007 (0.0006)
Control group mean	0.7638	0.5039	0.6693	0.6535	0.4567	0.6142	3.6614
Control group standard deviation	0.4264	0.5020	0.4723	0.4777	0.5001	0.4887	2.3611
Observations	666	666	666	666	666	666	666

Notes: Controls include state fixed effects, an index of social capital, percapita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

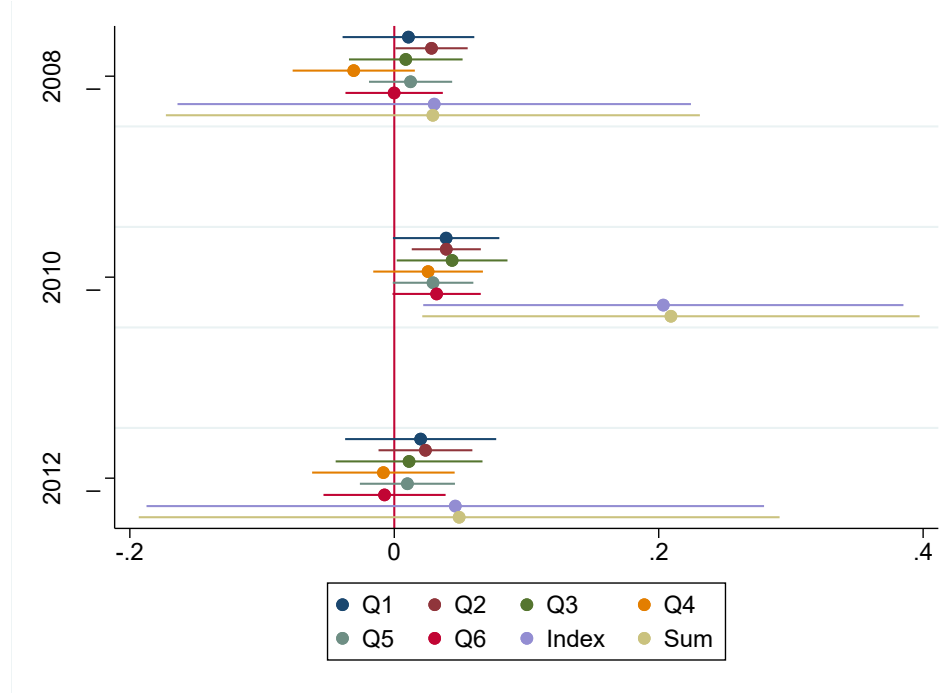


Figure 4.2: Event Analysis for the Full Sample

To provide evidence supporting the parallel trends assumption, I add two more waves of ENIGH data (2008 and 2010), and estimate the following equation:

$$Y_{itm} = \gamma T_m + \eta_t + \beta DID_{mt} + \delta X_{itm} + \epsilon_{itm} \quad (4.2)$$

This equation estimates difference-in-differences coefficients for all waves to test for parallel trends in the past. The thought behind this exercise is that evidence of parallel trends in the past supports the assumption of counterfactual parallel trends in the future. Even though this is not a direct test of the untestable assumption, it has become fairly popular in difference-in-differences papers.

Results of this estimation for the full and restricted samples are shown in Figures 4.2 and 4.3. The figures show that the difference-in-differences coefficients are not significant for any of the food security questions or the sum of them, except in the full population for 2010. Figure 4.2 shows statistically significant coefficients for most of the dependent variables in 2010. This may provide some additional evidence that looking for effects in the most food insecure households is a better idea than to compare the general population.

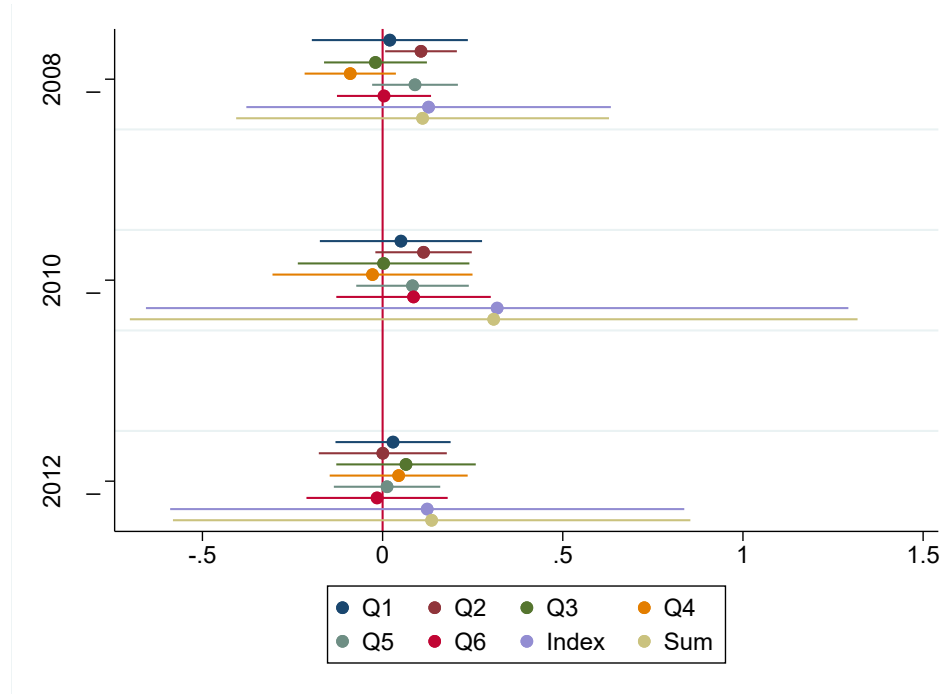


Figure 4.3: Event Analysis for the Restricted Sample

Figure 4.4 shows the comparison of the sum of all food security questions for treated and untreated households in time. Comparisons for the entire sample are shown in the left panel, while the restricted sample appears in the right panel. We can see that in both panels treatment groups have slightly higher sums on average (more food insecure), and travel quite close to the average of untreated households. However, in the left panel we do see the two lines becoming closer between 2010 and 2012, echoing what Figure 4.2 showed. The right panel shows trends that are closer together, almost fully overlapping, and switching orders between 2012 and 2014, reflecting the results from Table 4.7.

4.4.1 Effects on Food Expenditure

It is possible that the soup kitchen program did not only affect food security, but also assisted households in other spheres. One example would be how much people are spending on food items. Households may have substituted food prepared at home or bought elsewhere for the services of soup kitchens, which would slacken households' budget constraints, increasing their welfare. The ENIGH includes data on household expenditure on different goods. I estimate Equation 1 with a continuous treatment definition, using quarterly per-capita expenditures on food inside the home, outside the home, and the sum total. Table 4.8 shows the results for two sets of households: the restricted set within the most food insecure municipalities, and the unrestricted set.

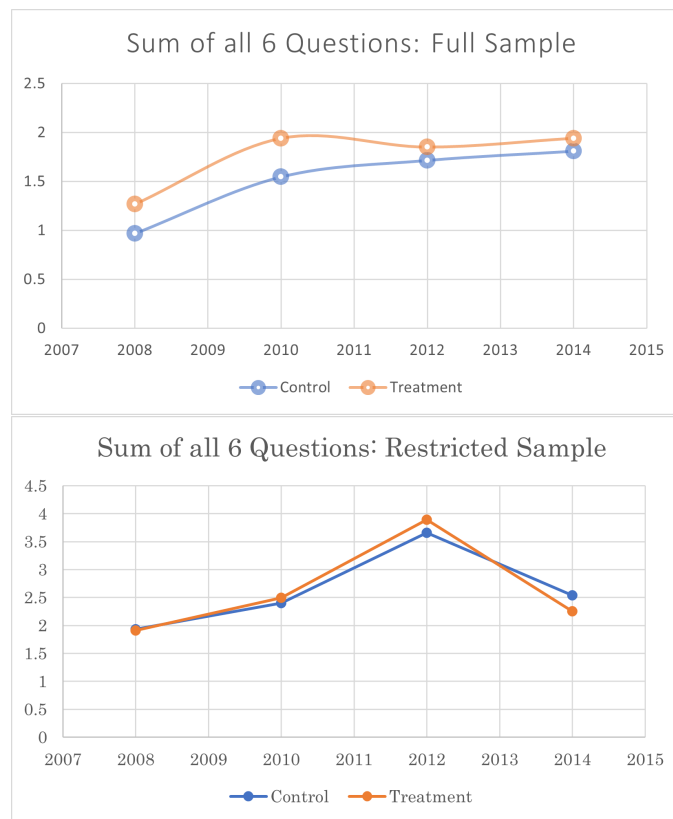


Figure 4.4: Trends for the Sum of Food Security Dummy Variables

Table 4.8: Estimation of Equation 1 with Continuous Treatment on Quarterly Per Capita Expenditure on Food, in Mexican Pesos

	(1)	(2)	(3)	(4)	(5)	(6)
	Outside the Home	Inside the Home	Total	Outside the Home	Inside the Home	Total
DID	-38.17*** (11.49)	-55.20 (46.84)	-93.83** (41.06)	1.404 (1.842)	-6.964** (3.389)	-5.530 (3.924)
After	182.8** (68.15)	368.7 (230.9)	552.9*** (169.3)	-8.563 (25.57)	116.9*** (40.83)	106.0** (50.21)
T=Number of SK	-0.606 (1.865)	11.55 (21.10)	10.97 (20.35)	-2.089 (2.418)	7.253** (3.206)	5.023 (4.427)
Control group mean	271.684	1540.943	1814.721	326.283	1717.232	2051.804
Control group standard deviation	656.848	1216.122	1523.688	743.517	1067.045	1427.768
Observations	350	350	350	6036	6036	6036

Notes: Controls include state fixed effects, an index of social capital, percapita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.8 shows that for households in the most ex-ante food insecure municipalities, all expenditure in food decreases. Since soup kitchens provided heavily subsidized meals, these results are consistent with households consuming these meals. It is interesting that the decrease in food expenditure seems to come from expenses both in and outside the home, perhaps suggesting that households substitute away from various forms of food access, and the soup kitchen is not acting as a substitute for a unique mechanism. The effect is smaller for the sample of all municipalities, as would be expected. It seems to be that even if the effect on food security was modest, households still benefited from the program in other margins. This highlights the arguments previously made in the literature that food security is a complicated phenomenon, and more income may not be sufficient to make an impact.

The fact that food expenditures decrease in this sample of municipalities, together with only modest effects on food security indicate that this slackening of the budget constraint is not sufficient to improve people's food security. One possible reason for this is that the food provided by the soup kitchens is simply substituting out an equivalent meal that households were consuming, and the additional cash is used for other purposes that households deem more pressing. Another possibility is that households cannot use this additional cash to improve their food security, perhaps because they do not have access to food items that would allow them to.

Furthermore, it is possible that there is a larger effect on expenditure than food security due to spillover effects of soup kitchens. For example, soup kitchens may have price effects where they are set up. Additionally, soup kitchens may be serving people, even in ex-ante food insecure municipalities, who are not the most food insecure. Thus, some people will be slackening their budget constraints, but not using that slack for food security because it is not a pressing issue to them. These potential explanations should be carefully studied, since they have very different implications for program design.

4.4.2 Political Alignment

Given the rules of the program, not only is community cooperation necessary for treatment, but some collaboration of local authorities is crucial to opening a soup kitchen. In particular, cooperation with the authorities is important for finding the physical space in which the Soup Kitchen is to be set up, and perhaps for navigating the funding application process. Using a triple differences specification, I explore the hypothesis that average treatment effects vary with the degree of cooperation between local authorities and the federal government.

There is some literature on government level collaboration that has shown that political party alignment is relevant for policy implementation. For example Durante and

Gutierrez find that neighboring municipalities that are politically aligned catch criminals more frequently and have fewer violent crimes. Brollo and Nannicini (2012) find that federal-municipal political alignment increases discretionary transfers of federal funds to municipalities.

This program was part of the flagship Crusade of the Mexican Federal Government, led at the time by Enrique Peña Nieto, from the Partido Revolucionario Institucional (PRI). It is conceivable that there was less cooperation between the Federally funded program and non-PRI-led municipal governments than PRI-led municipal governments, since political incentives were larger for PRI-led municipalities. This could be told as a story of local political sabotage of the Federal policy, or Federal sabotage of local administrations.

To explore this possibility, I coded an indicator variable for whether the PRI was governing each municipality when the Program was implemented (2013). I performed a triple difference estimation, and I present the results in Table 4.9. Results suggest that political alignment between the municipality and the Federal government has no impact on the distribution of the treatment effect. I also performed the analysis for state-level political alignment and the results remain the same. This is evidence that the lack of average effect cannot be attributed to a lack of cooperation of non-PRI mayors with the program.

The key feature of Table 4.9 is that neither the difference-in-differences coefficient nor the triple differences coefficient are statistically different from zero for any of the dependent variables. These regressions present more evidence of a muted mean effect of the program, and they reject the hypothesis of non-PRI municipal government sabotage to a federal program.

4.5 Conclusions

I estimated no municipal mean effect of the Soup Kitchen Program of the Mexican Crusade Against Hunger on any of six available measures of food security. However, when focusing only in the most food insecurity municipalities, I estimate a modest effect on the sum of all six food security questions. I find no evidence that political sabotage is responsible for the lack of average municipal effect.

Estimating effects of the soup kitchen program on food expenditures in and outside the home, I find that the program significantly decreases food expenditures for those in the most ex-ante food insecure municipalities. This result highlights the importance of looking at different spheres of beneficiaries' lives. While the effects of the soup kitchen program on food security may be small and focused only in the very tail of the distribution, the program seems to act also by slackening the budget constraints of some households. Even in this slackening does not translate into better food security, it is likely also welfare increasing.

Table 4.9: Estimation of Equation 3 with Binary Treatment on Measures of Food Security for the Restricted Sample

	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5	(6) Q6	(7) Sum
Triple Differences	0.029 (0.067)	-0.031 (0.041)	-0.051 (0.061)	0.017 (0.056)	-0.038 (0.039)	0.000 (0.052)	-0.074 (0.257)
After	0.053* (0.0291)	0.042 (0.0257)	0.023 (0.03)	0.014 (0.027)	0.065*** (0.0217)	0.038 (0.0255)	0.236* (0.128)
T=SK Open	-0.001 (0.001)	0.001 (0.001)	-0.002 (0.001)	-0.002 (0.001)	0.000 (0.002)	0.000 (0.002)	-0.003 (0.008)
DID	-0.048 (0.031)	-0.024 (0.029)	-0.021 (0.040)	-0.006 (0.033)	-0.022 (0.026)	0.002 (0.037)	-0.120 (0.175)
A*PRI	-0.044 (0.046)	-0.002 (0.033)	0.051 (0.041)	0.005 (0.046)	-0.037 (0.029)	-0.013 (0.040)	-0.040 (0.192)
T*PRI	-0.020 (0.052)	0.021 (0.034)	0.063 (0.043)	-0.018 (0.044)	0.014 (0.034)	-0.003 (0.037)	0.057 (0.194)
PRI	0.046 (0.064)	-0.015 (0.059)	-0.059 (0.061)	-0.023 (0.067)	0.030 (0.065)	0.018 (0.057)	-0.002 (0.321)
Observations	5,701	5,701	5,701	5,701	5,701	5,701	5,701

Notes: Controls include state fixed effects, an index of social capital, per capita household income, gender of the household head, number of children in the household. Standard errors, clustered by municipality, are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

This slackening on the budget constraint could be coming from households buying less food because they receive it at the soup kitchen, or from general equilibrium effects of soup kitchens in the food provision market. In that sense, it is possible that the effects I estimate in this chapter include large spillovers to households who do not attend soup kitchens but still benefit. These spillovers should also be studied, as they may be especially large in specific types of communities.

Out of the three initially mentioned channels through which the program could be failing to improve mean food security, this chapter presents evidence against the hypothesis of lack of implementation cooperation, and evidence against bad targeting at the municipality level. It is possible, however, that households that are the most food insecure within a municipality do not make use of the program. With the current data, it is hard to infer actual participation in the program, but this should be further pursued. Other possible explanations that should be further studied include crowding out, and insufficient quality of food provided.

Chapter 5

Conclusion

Different and complex barriers exist to widespread improvements in the standard of living conditions. Identifying them and systematically studying them is imperative to improve the lives of those who have not benefited as much from the rise in living conditions the world has seen in the last century.

As governments and other organizations design policies to reduce these barriers, the complexities of the different phenomena must be taken into account. In this dissertation I study barriers in three different margins: gender and the labor market, subjective belief elicitation, and food security.

In all three chapters, the importance of understanding the complexities of barriers in different aspects of peoples lives cannot be understated. Chapter 2 shows that women's human capital investment is constrained by the work-family tradeoffs they face. To achieve gender equality, differences in these tradeoffs between men and women need to be addressed, as they impact more than just labor supply. Chapter 3 shows that when one wants to study complex and important phenomena, such as the effects of subjective expectations in behavior, one must be careful to measure these without bias. Chapter 4 provides more evidence that food security is a complex phenomenon, and that seemingly intuitive and simple solutions may not be enough to increase it. It also emphasizes that sometimes policymakers set goals and objectives that may not be the same as the priorities of beneficiaries. If we wish to improve peoples' lives, we need to understand what are the truly important barriers they face, from their perspective.

References

- Alberto Abadie. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. page 44, 2019.
- Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, March 2003. doi: 10.1257/000282803321455188. URL <http://www.aeaweb.org/articles?id=10.1257/000282803321455188>.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505, June 2010. ISSN 0162-1459, 1537-274X. doi: 10.1198/jasa.2009.ap08746.
- Claire Adida, Daniel Posner, and Amanda Lee Robinson. Interviewer Coethnicity Effects in African Survey Data. Technical report, November 2011.
- Claire L. Adida, Karen E. Ferree, Daniel N. Posner, and Amanda Lea Robinson. Who’s Asking? Interviewer Coethnicity Effects in African Survey Data. *Comparative Political Studies*, 49(12):1630–1660, October 2016. ISSN 0010-4140. doi: 10.1177/0010414016633487.
- Emma Aguila, Abril Borges, Cielo Margot Castillejos, Ashley Pierson, and Beverly A. Weidmer. Mortality Expectations of Older Mexicans: Development and Testing of Survey Measures. https://www.rand.org/pubs/technical_reports/TR1288z6.html, 2014.
- Mohammad Alhaj, Alemayehu Amberbir, Emmanuel Singogo, Victor Banda, Monique van Lettow, Alfred Matengeni, Gift Kawalazira, Joe Theu, Megh R Jagriti, Adrienne K Chan, and Joep J van Oosterhout. Retention on antiretroviral therapy during Universal Test and Treat implementation in Zomba district, Malawi: A retrospective cohort study. *Journal of the International AIDS Society*, 22(2), February 2019. ISSN 1758-2652. doi: 10.1002/jia2.25239.
- Barbara A. Anderson, Brian D. Silver, and Paul R. Abramson. The Effects of the Race of the Interviewer on Race-related Attitudes of Black Respondents in SRC/CPS National

- Election Studies. *Public Opinion Quarterly*, 52(3):289–324, January 1988. ISSN 0033-362X. doi: 10.1086/269108.
- Anna V. Andreenkova. How to Choose Interview Language in Different Countries. In Timothy P. Johnson, Beth-Ellen Pennell, Ineke AL Stoop, and Brita Dorer, editors, *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, Wiley Series in Survey Methodology. John Wiley & Sons, Hoboken, NJ, 2018. ISBN 978-1-118-88496-6 978-1-118-88501-7.
- Philip Anglewicz and Jesman Chintsanya. Disclosure of HIV status between spouses in rural Malawi. *AIDS Care*, 23(8):998–1005, August 2011. ISSN 0954-0121, 1360-0451. doi: 10.1080/09540121.2010.542130.
- Philip Anglewicz and Shelley Clark. The effect of marriage and HIV risks on condom use acceptability in rural Malawi. *Social Science & Medicine*, 97:29–40, November 2013. ISSN 0277-9536. doi: 10.1016/j.socscimed.2013.06.024.
- Philip Anglewicz and Hans-Peter Kohler. Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi. *Demographic research*, 20(6):65–96, 2009. doi: 10.4054/DemRes.2009.20.6.
- Philip Anglewicz and Georges Reniers. HIV Status, Gender, and Marriage Dynamics among Adults in Rural Malawi. *Studies in Family Planning*, 45(4):415–428, 2014. ISSN 1728-4465. doi: 10.1111/j.1728-4465.2014.00005.x.
- N. Angotti, K. Y. Dionne, and L. Gaydos. An offer you can’t refuse? Provider-initiated HIV testing in antenatal clinics in rural Malawi. *Health Policy and Planning*, 26(4):307–315, 2011.
- Animal-Politico. Los comedores comunitarios entregan alimentos, pero no acaban con el hambre; Sedesol alista cambios. 2016.
- W. S. Aquilino, D. L. Wright, and A. J. Supple. Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, 35(6-8):845–867, 2000 May-Jun. ISSN 1082-6084.
- Orazio Attanasio, Costas Meghir, and Marcos Vera-Hernández. Elicitation, validation, and use of probability distributions of future income in developing countries. Working Paper, 2005.
- Orazio P Attanasio. Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review*, 99(2):87–92, April 2009. ISSN 0002-8282. doi: 10.1257/aer.99.2.87.

- Martha Bailey, Tanya Byker, Elena Patel, and Shanthi Ramnath. The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data. Technical Report w26416, National Bureau of Economic Research, Cambridge, MA, October 2019.
- Ann P. Bartel, Maya Rossin-Slater, Christopher J. Ruhm, Jenna Stearns, and Jane Waldfogel. Paid Family Leave, Fathers' Leave-Taking, and Leave-Sharing in Dual-Earner Households. *Journal of Policy Analysis and Management*, 37:10–37, 2018. ISSN 1520-6688. doi: 10.1002/pam.22030.
- Charles Baum and Christopher Ruhm. The effects of paid family leave in california on labor market outcomes. NBER Working Papers 19741, National Bureau of Economic Research, Inc, 2013. URL <https://EconPapers.repec.org/RePEc:nbr:nberwo:19741>.
- Charles L. Baum and Christopher J. Ruhm. The Effects of Paid Family Leave in California on Labor Market Outcomes. *Journal of Policy Analysis and Management*, 35(2):333–356, 2016. ISSN 1520-6688. doi: 10.1002/pam.21894. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.21894>.
- Stan Becker, Kale Feyisetan, and Paulina Makinwa-Adebusoye. The effect of the sex of interviewers on the quality of data in a Nigerian family planning questionnaire. <https://link.galegroup.com/apps/doc/A17474999/AONE?sid=lms>, July 1995.
- Kelly Bedard and Maya Rossin-Slater. The Economic and Social Impacts of Paid Family Leave in California: Report for the California Employment Development Department. page 71, 2016.
- Robert F. Belli, William L. Shay, and Frank P. Stafford. Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly*, 65(1):45–74, March 2001. ISSN 0033-362X, 1537-5331. doi: 10.1086/320037.
- Lori Bennear, Alessandro Tarozzi, Alexander Pfaff, Soumya Balasubramanya, Kazi Matin Ahmed, and Alexander van Geen. Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in Bangladesh. *Journal of Environmental Economics and Management*, 65(2):225–240, March 2013. ISSN 0095-0696. doi: 10.1016/j.jeem.2012.07.006.
- S. Bignami-Van Assche, L. W. Chao, P. Anglewicz, D. Chilongozi, and A. Bula. The validity of self-reported likelihood of HIV infection among the general population in rural Malawi. *Sexually transmitted infections*, 83(1):35–40, 2007.

- Lisa Blaydes and Rachel M. Gillum. Religiosity-of-Interviewer Effects: Assessing the Impact of Veiled Enumerators on Survey Response in Egypt. *Politics and Religion*, 6(3):459–482, September 2013. ISSN 1755-0483, 1755-0491. doi: 10.1017/S1755048312000557.
- Fernanda Brollo and Tommaso Nannicini. Tying Your Enemy’s Hands in Close Races: The Politics of Federal Transfers in Brazil. *The American Political Science Review*, 106(4): 742–761, 2012. ISSN 0003-0554.
- B. Burström and P. Fredlund. Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology & Community Health*, 55(11):836–840, November 2001. ISSN 0143-005X, 1470-2738. doi: 10.1136/jech.55.11.836.
- Tanya S. Byker. Paid Parental Leave Laws in the United States: Does Short-Duration Leave Affect Women’s Labor-Force Attachment? *American Economic Review*, 106(5): 242–246, May 2016. ISSN 0002-8282. doi: 10.1257/aer.p20161118.
- Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326, November 2014. ISSN 00129682. doi: 10.3982/ECTA11757.
- Sebastian Calonico, Matias D. Cattaneo, and Rocío Titiunik. Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association*, 110(512):1753–1769, October 2015. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2015.1017578.
- Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. Rdrobust: Software for Regression-discontinuity Designs. *The Stata Journal: Promoting communications on statistics and Stata*, 17(2):372–404, June 2017. ISSN 1536-867X, 1536-8734. doi: 10.1177/1536867X1701700208.
- Pedro Carneiro, Katrine V. Løken, and Kjell G. Salvanes. A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children. *Journal of Political Economy*, 123(2): 365–412, April 2015a. ISSN 0022-3808. doi: 10.1086/679627.
- Pedro Carneiro, Katrine V. Løken, and Kjell G. Salvanes. A flying start? maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, 123(2):365–412, 2015b. doi: 10.1086/679627. URL <https://doi.org/10.1086/679627>.
- Moses Vernonxious Madalitso Chamba. *Evaluation Study on Implementation of Life Skills Programme in Public Secondary Schools in Malawi: South Eastern Educational Division (SEED)*. PhD thesis, University of Malawi College of Medicine, May 2009.

- Gretchen B. Chapman and Eric J. Johnson. Incorporating the Irrelevant: Anchors in Judgments of Belief and Value. In Thomas Gilovich, Dale Griffin, and Daniel Kahneman, editors, *Heuristics and Biases*, pages 120–138. Cambridge University Press, first edition, July 2002. ISBN 978-0-521-79260-8 978-0-511-80809-8 978-0-521-79679-8. doi: 10.1017/CBO9780511808098.008.
- Jobiba Chinkhumba, Susan Godlonton, and Rebecca Thornton. Demand for medical male circumcision. *American Economic Journal: Applied Economics*, 6(2):152–177, 2014.
- Grames Chirwa and Devika Naidoo. Structural and social constraints in the teaching of Life Skills for HIV/AIDS prevention in Malawi primary schools. *South African Journal of Childhood Education*, 4(1):19, 2014.
- Yating Chuang and Laura Schechter. Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, 117:151–170, November 2015. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2015.07.008.
- Jacobus Cilliers, Oeindrila Dube, and Bilal Siddiqi. The white-man effect: How foreigner presence affects behavior in experiments. *Journal of Economic Behavior & Organization*, 118:397–414, October 2015. ISSN 0167-2681. doi: 10.1016/j.jebo.2015.03.015.
- CONEVAL. Estructura alcance instrumentos CNCH. Technical report, 2014.
- CONEVAL. Estudio Exploratorio de los Comedores Comunitarios SEDESOL. Technical report, CONEVAL, 2015.
- Patrick R. Cotter, Jeffrey Cohen, and Philip B. Coulter. Race-of-Interviewer Effects in Telephone Interviews. *Public Opinion Quarterly*, 46(2):278–284, January 1982. ISSN 0033-362X. doi: 10.1086/268719.
- Darren W. Davis. Nonrandom Measurement Error and Race of Interviewer Effects Among African Americans. *The Public Opinion Quarterly*, 61(1):183–207, 1997. ISSN 0033-362X.
- Suresh de Mel, David McKenzie, and Christopher Woodruff. Returns to Capital in Microenterprises: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 123(4):1329–1372, November 2008. ISSN 0033-5533. doi: 10.1162/qjec.2008.123.4.1329.
- Jonathan de Quidt, Johannes Haushofer, and Christopher Roth. Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11):3266–3302, November 2018. ISSN 0002-8282. doi: 10.1257/aer.20171330.

- Adeline Delavande. Pill, Patch, or Shot? Subjective Expectations and Birth Control Choice. *International Economic Review*, 49(3):999–1042, 2008. ISSN 0020-6598.
- Adeline Delavande. Probabilistic Expectations in Developing Countries. *Annual Review of Economics*, 6(1):1–20, August 2014. ISSN 1941-1383. doi: 10.1146/annurev-economics-072413-105148.
- Adeline Delavande and Hans-Peter Kohler. Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic research*, 20:817–874, June 2009. ISSN 1435-9871. doi: 10.4054/DemRes.2009.20.31.
- Adeline Delavande and Hans-Peter Kohler. The Impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi. *Demography*, 49(3):1011–1036, August 2012. ISSN 1533-7790. doi: 10.1007/s13524-012-0119-7.
- Adeline Delavande and Hans-Peter Kohler. HIV/AIDS-related Expectations and Risky Sexual Behaviour in Malawi. *The Review of Economic Studies*, 83(1):118–164, January 2016. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdv028.
- Adeline Delavande and Susann Rohwedder. Differential Survival in Europe and the United States: Estimates Based on Subjective Probabilities of Survival. *Demography*, 48(4): 1377–1400, November 2011. ISSN 1533-7790. doi: 10.1007/s13524-011-0066-8.
- Adeline Delavande, Xavier Giné, and David McKenzie. Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, 94(2):151–163, March 2011a. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2010.01.008.
- Adeline Delavande, Xavier Giné, and David McKenzie. Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, 94(2):151–163, March 2011b. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2010.01.008.
- Adeline Delavande, Jinkook Lee, and Seetha Menon. Eliciting Survival Expectations of the Elderly in Low-Income Countries: Evidence From India. *Demography*, 54(2):673–699, April 2017. ISSN 1533-7790. doi: 10.1007/s13524-017-0560-8.
- Michele Di Maio and Nathan Fiala. Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect. *The World Bank Economic Review*, in press, 2019. doi: 10.1093/wber/lhy024.

- Kim Yi Dionne. The politics of local research production: Surveying in a context of ethnic competition. *Politics, Groups, and Identities*, 2(3):459–480, July 2014. ISSN 2156-5503. doi: 10.1080/21565503.2014.930691.
- Brigitte Dormont, Anne-Laure Samson, Marc Fleurbaey, Stéphane Luchini, and Erik Schokkaert. Individual Uncertainty About Longevity. *Demography*, 55(5):1829–1854, October 2018. ISSN 1533-7790. doi: 10.1007/s13524-018-0713-4.
- Ruben Durante and Emilio Gutierrez. Fighting Crime with a Little Help from my Friends: Party Affiliation, Inter-jurisdictional Cooperation and Crime in Mexico. page 32.
- Christian Dustmann and Uta Schönberg. Expansions in maternity leave coverage and children’s long-term outcomes. *American Economic Journal: Applied Economics*, 4(3):190–224, July 2012. doi: 10.1257/app.4.3.190. URL <http://www.aeaweb.org/articles?id=10.1257/app.4.3.190>.
- Todd E. Elder. The Predictive Validity of Subjective Mortality Expectations: Evidence From the Health and Retirement Study. *Demography*, 50(2):569–589, April 2013. ISSN 1533-7790. doi: 10.1007/s13524-012-0164-2.
- Theresa M. Fedor, Hans-Peter Kohler, and Jere R. Behrman. The Impact of Married Individuals Learning HIV Status in Malawi: Divorce, Number of Sexual Partners, and Condom Use With Spouses. *Demography*, 52(1):259–280, January 2015. ISSN 0070-3370, 1533-7790. doi: 10.1007/s13524-014-0364-z.
- Steven E. Finkel, Thomas M. Guterbock, and Marian J. Borg. RACE-OF-INTERVIEWER EFFECTS IN A PREELECTION POLL VIRGINIA 1989. *Public Opinion Quarterly*, 55(3):313–330, January 1991. ISSN 0033-362X. doi: 10.1086/269264.
- Lars Frison and Stuart J. Pocock. Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in medicine*, 11(13):1685–1704, September 1992. ISSN 0277-6715.
- Melanie Gallant and Eleanor Maticka-Tyndale. School-based HIV prevention programmes for African youth. *Social Science & Medicine*, 58(7):1337–1351, April 2004. ISSN 0277-9536. doi: 10.1016/S0277-9536(03)00331-9.
- Susan Godlonton, Alister Munthali, and Rebecca Thornton. Responding to Risk: Circumcision, Information, and HIV Prevention. *Review of Economics and Statistics*, page in press, 2015.

- Monica J. Grant. Children's school participation and HIV/AIDS in rural Malawi:: The role of parental knowledge and perceptions. *Demographic Research*, 19:1603–1634, September 2008. ISSN 1435-9871. doi: 10.4054/DemRes.2008.19.45.
- Monica J. Grant and Erica Soler-Hampejsek. HIV Risk Perceptions, the Transition to Marriage, and Divorce in Southern Malawi. *Studies in Family Planning*, 45(3):315–337, 2014. ISSN 1728-4465. doi: 10.1111/j.1728-4465.2014.00394.x.
- NIMH Collaborative HIV/STD Prevention Trial Group. The feasibility of audio computer-assisted self-interviewing in international settings:. *AIDS*, 21(Suppl 2):S49–S58, April 2007. ISSN 0269-9370. doi: 10.1097/01.aids.0000266457.11020.f0.
- Craig Gundersen and Brent Kreider. Food Stamps and Food Insecurity What Can Be Learned in the Presence of Nonclassical Measurement Error? *Journal of Human Resources*, 43(2):352–382, March 2008. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.43.2.352.
- Craig Gundersen and Brent Kreider. Bounding the effects of food insecurity on children's health outcomes. *Journal of Health Economics*, 28(5):971–983, September 2009. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2009.06.012.
- Craig Gundersen, Brent Kreider, and John Pepper. The impact of the National School Lunch Program on child health: A nonparametric bounds analysis. *Journal of Econometrics*, 166(1):79–91, January 2012. ISSN 0304-4076. doi: 10.1016/j.jeconom.2011.06.007.
- E. A. Hahn, D. Cellal, D. G. Dobrez, G. Shiimoto, S. G. Taylor, A. G. Galvez, P. Diaz, V. Valenzuela, H. L. Chiang, S. Khan, S. A. Hudgens, and H. Du. Quality of life assessment for low literacy Latinos: A new multimedia program for self-administration. *The Journal of oncology management : the official journal of the American College of Oncology Administrators*, 12(5):9–12, 2003. ISSN 1061-9364.
- Wen-Jui Han, Christopher Ruhm, and Jane Waldfogel. Parental leave policies and parents' employment and leave-taking. *Journal of Policy Analysis and Management*, 28(1):29–54, 2009. ISSN 1520-6688. doi: 10.1002/pam.20398. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.20398>.
- Stéphane Helleringer and Hans-Peter Kohler. Social networks, perceptions of risk, and changing attitudes towards HIV/AIDS: New evidence from a longitudinal study using fixed-effects analysis. *Population Studies*, 59(3):265–282, November 2005. ISSN 0032-4728. doi: 10.1080/00324720500212230.

- Paul C. Hewett, Annabel S. Erulkar, Barbara S. Mensch, and Council Kenya. The feasibility of computer-assisted survey interviewing in Africa: Experience from two rural districts in Kenya. *Social Science Computer Review*, 2004.
- Larry L. Howard. Transitions between food insecurity and food security predict children's social skill development during elementary school. *British Journal of Nutrition*, 105(12): 1852–1860, June 2011. ISSN 1475-2662, 0007-1145. doi: 10.1017/S0007114510005623.
- Péter Hudomiet, Gábor Kézdi, and Robert J. Willis. Stock Market Crash and Expectations of American Households. *Journal of Applied Econometrics*, 26(3):393–415, 2011. ISSN 1514-0326. doi: 10.1002/jae.1226.
- Michael D. Hurd and Kathleen McGarry. Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study. *The Journal of Human Resources*, 30: S268–S292, 1995. ISSN 0022-166X. doi: 10.2307/146285.
- Ellen L. Idler and Yael Benyamini. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*, 38(1):21–37, 1997. ISSN 0022-1465. doi: 10.2307/2955359.
- Guido W. Imbens and Donald B. Rubin. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. /core/books/causal-inference-for-statistics-social-and-biomedical-sciences/71126BE90C58F1A431FE9B2DD07938AB, April 2015.
- Jennifer A. Johnson-Hanks. Vital Conjunctures Revisited. In Philip Kreager and Astrid Bochow, editors, *Fertility, Conjuncture, Difference: Anthropological Approaches to the Heterogeneity of Modern Fertility Declines*, pages 324–338. Berghahn Books, New York, first edition, September 2017. ISBN 978-1-78533-605-8 978-1-78533-604-1. doi: 10.2307/j.ctvw04c56.
- Frank Kadyoma, Mumbi Mbugua, Agnes Ndung'u, and Edward Omae. *Malawi Junior Secondary Life Skills: Form 2*. Longhorn Publishers, Nairobi, Kenya, 2012.
- Amy Kaler. "My girlfriends could fill a yanu-yanu bus": Rural Malawian men's claims about their own serostatus. *Demographic Research*, Special Collection(1), 2003.
- Amy Kaler and Susan Watkins. Asking God about the date you will die: HIV testing as a zone of uncertainty in rural Malawi. *Demographic Research*, 23, 2010.
- Amy Kaler, Susan Cotts Watkins, and Nicole Angotti. Making meaning in the time of AIDS: Longitudinal narratives from the Malawi Journals Project. *African Journal of AIDS Research*, 14(4):303–314, October 2015. ISSN 1608-5906. doi: 10.2989/16085906.2015.1084342.

- Jason Kerwin. Scared Straight or Scared to Death? The Effect of Risk Beliefs on Risky Behaviors. Working Paper, University of Minnesota, 2020.
- Jason T. Kerwin, Rebecca L. Thornton, Sallie M. Foley, Jobiba Chinkhumba, and Alinafe Chibwana. *Situational Analysis of Sexual Behaviors and Alternative Safer Sex Strategies In-Depth Interview Dataset*. University of Michigan and University of Malawi College of Medicine, 2011.
- Pinar Keskin, Gauri Kartini Shastry, and Helen Willis. Water Quality Awareness and Breastfeeding: Evidence of Health Behavior Change in Bangladesh. *Review of Economics and Statistics*, August 2016. ISSN 0034-6535. doi: 10.1162/REST_a_00626.
- Jacob Alex Klerman and Arleen Leibowitz. Labor Supply Effects of State Maternity Leave Legislation:. <https://www.rand.org/pubs/reprints/RP715.html>, 1998.
- Jochen Kluge and Marcus Tamm. Parental leave regulations, mothers’ labor force attachment and fathers’ childcare involvement: Evidence from a natural experiment. *Journal of Population Economics*, 26(3):983–1005, July 2013. ISSN 1432-1475. doi: 10.1007/s00148-012-0404-1.
- Hans-Peter Kohler, Jere R. Behrman, and Susan C. Watkins. Social networks and HIV/AIDS risk perceptions. *Demography*, 44(1):1–33, February 2007. ISSN 1533-7790. doi: 10.1353/dem.2007.0006.
- Brent Kreider, John V. Pepper, Craig Gundersen, and Dean Jolliffe. Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported. *Journal of the American Statistical Association*, 107(499):958–975, September 2012. ISSN 0162-1459. doi: 10.1080/01621459.2012.682828.
- Brent Kreider, John V. Pepper, and Manan Roy. Identifying the Effects of WIC on Food Insecurity Among Infants and Children. *Southern Economic Journal*, 82(4):1106–1122, 2016. ISSN 2325-8012. doi: 10.1002/soej.12078.
- Michael Kremer and Edward Miguel. The Illusion of Sustainability. *The Quarterly Journal of Economics*, 122(3):1007–1065, 2007. doi: 10.1162/qjec.122.3.1007.
- Ilyana Kuziemko, Jessica Pan, Jenny Shen, and Ebonya Washington. The mommy effect: Do women anticipate the employment effects of motherhood? Technical Report w24740, National Bureau of Economic Research, Cambridge, MA, June 2018.
- Gianmarco León and Edward Miguel. Risky Transportation Choices and the Value of a Statistical Life. *American Economic Journal: Applied Economics*, 9(1):202–228, January 2017. ISSN 1945-7782. doi: 10.1257/app.20160140.

- Laurent Lequien. The Impact of Parental Leave Duration on Later Wages. *Annals of Economics and Statistics*, (107/108):267–285, 2012. ISSN 2115-4430. doi: 10.2307/23646579.
- Lee A. Lillard and Robert J. Willis. Cognition and Wealth: The Importance of Probabilistic Thinking. SSRN Scholarly Paper ID 1084559, Social Science Research Network, Rochester, NY, June 2001.
- Nancy Luke, Shelley Clark, and Eliya M. Zulu. The Relationship History Calendar: Improving the scope and quality of data on youth sexual behavior. *Demography*, 48(3):1151–1176, 2011.
- K. Mac Dougall, Y. Beyene, and R. D. Nachtigall. Age shock: Misperceptions of the impact of age on fertility before and after IVF in women who conceived after age 40. *Human Reproduction*, 28(2):350–356, February 2013. ISSN 0268-1161. doi: 10.1093/humrep/des409.
- Mary Mahony. Fertility Knowledge Among US Women Aged 25–35: Insights from a New Generation. Technical report, Medaxial, October 2011.
- Charles F. Manski. Measuring Expectations. *Econometrica*, 72(5):1329–1376, 2004. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2004.00537.x.
- Katie S Martin, Beatrice L Rogers, John T Cook, and Hugh M Joseph. Social capital is associated with decreased risk of hunger. *Social Science & Medicine*, 58(12):2645–2654, June 2004. ISSN 0277-9536. doi: 10.1016/j.socscimed.2003.09.026.
- Daniel Maxwell, Bapu Vaitla, and Jennifer Coates. How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. *Food Policy*, 47:107–116, August 2014. ISSN 0306-9192. doi: 10.1016/j.foodpol.2014.04.003.
- Susan McCombie and John Anarfi. The influence of sex of interviewer on the results of an aids survey in ghana. *Human Organization*, 61:51–57, 03 2002. doi: 10.17730/humo.61.1.em6l865y3v9y7y2l.
- David McKenzie, John Gibson, and Steven Stillman. A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics*, 102:116–127, May 2013. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2012.01.001.
- David J. McKenzie, John Gibson, and Steven Stillman. How Important is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration. SSRN Scholarly Paper ID 898591, Social Science Research Network, Rochester, NY, April 2006.

- Barbara S. Mensch, Paul C. Hewett, and Annabel S. Erulkar. The Reporting of Sensitive Behavior by Adolescents: A Methodological Experiment in Kenya. *Demography*, 40(2): 247–268, 2003. ISSN 0070-3370. doi: 10.2307/3180800.
- Barbara S. Mensch, Paul C. Hewett, Richard Gregory, and Stephane Helleringer. Sexual Behavior and STI/HIV Status Among Adolescents in Rural Malawi: An Evaluation of the Effect of Interview Mode on Reporting. *Studies in Family Planning*, 39(4):321–334, 2008. ISSN 1728-4465. doi: 10.1111/j.1728-4465.2008.00178.x.
- Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Technical report, IPUMS, Minneapolis, MN, 2019.
- Tamara Mousa and Jeanne Freeland-Graves. Food security of food recipients of a food pantry and soup kitchen. *Public Health Nutrition*, 22:1–10, February 2019. doi: 10.1017/S1368980018003658.
- Maximilian W Mueller, Joan Hamory Hicks, Jennifer Johnson-Hanks, and Edward Miguel. The Illusion of Stable Preferences over Major Life Decisions. Working Paper 25844, National Bureau of Economic Research, May 2019.
- A. C Munthali and E. M Zulu. The Timing and role of Initiation Rites in Preparing Young People for Adolescence and Responsible Sexual and Reproductive Behaviour in Malawi. *African Journal of Reproductive Health*, 11(3):150, 2007.
- Thomas Mussweiler, Fritz Strack, and Tim Pfeiffer. Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and Social Psychology Bulletin*, 26(9):1142–1150, November 2000. ISSN 0146-1672. doi: 10.1177/01461672002611010.
- National Statistical Office - NSO/Malawi and ICF. Malawi Demographic and Health Survey. Technical report, NSO and ICF, Zomba, Malawi, February 2017.
- Ben R. Newell and David R. Shanks. Prime Numbers: Anchoring and its Implications for Theories of Behavior Priming. *Social Cognition*, 32(Supplement):88–108, June 2014. ISSN 0278-016X. doi: 10.1521/soco.2014.32.supp.88.
- Martin T. Orne. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11): 776–783, 1962. ISSN 1935-990X(Electronic),0003-066X(Print). doi: 10.1037/h0043424.
- Jessica E Pac, Ann P Bartel, Christopher J Ruhm, and Jane Waldfogel. Paid Family Leave and Breastfeeding: Evidence from California. Working Paper 25784, National Bureau of Economic Research, April 2019.

- Maria Perozek. Using subjective expectations to forecast longevity: Do survey respondents know something we don't know? *Demography*, 45(1):95–113, February 2008. ISSN 1533-7790. doi: 10.1353/dem.2008.0010.
- Rukmini Potdar and Michael A. Koenig. Does Audio-CASI Improve Reports of Risky Behavior? Evidence from a Randomized Field Trial Among Young Urban Men in India. *Studies in Family Planning*, 36(2):107–116, 2005. ISSN 1728-4465. doi: 10.1111/j.1728-4465.2005.00048.x.
- Stephen D. Reese, Wayne A. Danielson, Pamela J. Shoemaker, Tsan-Kuo Chang, and Huei-Ling Hsu. Ethnicity-of-Interviewer Effects Among Mexican-Americans and Anglos. *Public Opinion Quarterly*, 50(4):563–572, January 1986. ISSN 0033-362X. doi: 10.1086/269004.
- Maya Rossin. The effects of maternity leave on children's birth and infant health outcomes in the United States. *Journal of Health Economics*, 30(2):221–239, March 2011. URL <https://ideas.repec.org/a/eee/jhecon/v30y2011i2p221-239.html>.
- Maya Rossin-Slater. Maternity and Family Leave Policy. Working Paper 23069, National Bureau of Economic Research, January 2017.
- Maya Rossin-Slater, Christopher J. Ruhm, and Jane Waldfogel. The effects of california's paid family leave program on mothers' leave-taking and subsequent labor market outcomes. Working Paper 17715, National Bureau of Economic Research, December 2011. URL <http://www.nber.org/papers/w17715>.
- Maya Rossin-Slater, Christopher J. Ruhm, and Jane Waldfogel. The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes. *Journal of Policy Analysis and Management*, 32(2):224–245, 2013. ISSN 1520-6688. doi: 10.1002/pam.21676. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.21676>.
- P Rumakom, P Guest, W Chinvarasopak, W Utarmat, and J Sontanakanit. Obtaining accurate responses to sensitive questions among Thai students: A comparison of two data collection techniques. In S Jejeebhoy, I Shah, and S Thapa, editors, *Sex Without Consent*, pages 318–332. Zed Books, London, 2005.
- SEDESOL. Lineamientos Especificos para el Programa de Comedores Comunitarios, en el Marco de la Cruzada Nacional Contra el Hambre, 2013.
- SEDESOL. Programa Nacional Mexico Sin Hambre 2014-2018. Logros 2014. Technical report, SEDESOL, 2015.

- Gil Shapira. How Subjective Beliefs about HIV Infection Affect Life-Cycle Fertility: Evidence from Rural Malawi. *Journal of Human Resources*, 52(3):680–718, 2017. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.52.3.0314-6263R1.
- Maheshwor Shrestha. Get Rich or Die Tryin’: Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal. *The World Bank Economic Review*, 34(1):1–27, February 2020. ISSN 0258-6770, 1564-698X. doi: 10.1093/wber/lhz023.
- Kirsten P. Smith and Susan Cotts Watkins. Perceptions of risk and strategies for prevention: Responses to HIV/AIDS in rural Malawi. *Social Science & Medicine*, 60(3):649–660, February 2005. ISSN 02779536. doi: 10.1016/j.socscimed.2004.06.009.
- Erica Soler-Hampejsek, Monica J. Grant, Barbara S. Mensch, Paul C. Hewett, and Johanna Rankin. The Effect of School Status and Academic Skills on the Reporting of Premarital Sexual Behavior: Evidence From a Longitudinal Study in Rural Malawi. *Journal of Adolescent Health*, 53(2):228–234, August 2013. ISSN 1054-139X. doi: 10.1016/j.jadohealth.2013.03.008.
- Jenna Stearns. The effects of paid maternity leave: Evidence from Temporary Disability Insurance. *Journal of Health Economics*, 43:85–102, September 2015. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2015.04.005.
- Rebecca L. Thornton. The Demand for, and Impact of, Learning HIV Status. *American Economic Review*, 98(5):1829–1863, 2008.
- Emilia Tjernström, Travis J Lybbert, Rachel Frattarola Hernández, and Juan Sebastian Correa. Learning by (virtually) doing: Experimentation and belief updating in small-holder agriculture. Working Paper, April 2019.
- Jenny Trinitapoli. Religious teachings and influences on the ABCs of HIV prevention in Malawi. *Social Science & Medicine*, 69(2):199–209, July 2009. ISSN 0277-9536. doi: 10.1016/j.socscimed.2009.04.018.
- Jenny Trinitapoli and Mark D. Regnerus. Religion and HIV Risk Behaviors Among Married Men: Initial Results from a Study in Rural Sub-Saharan Africa. *Journal for the Scientific Study of Religion*, 45(4):505–528, 2006. ISSN 1468-5906. doi: 10.1111/j.1468-5906.2006.00325.x.
- Jenny Trinitapoli and Sara Yeatman. Uncertainty and Fertility in a Generalized AIDS Epidemic. *American Sociological Review*, 76(6):935–954, December 2011. ISSN 0003-1224. doi: 10.1177/0003122411427672.

- Jenny Trinitapoli and Sara Yeatman. The Flexibility of Fertility Preferences in a Context of Uncertainty: Flexibility of Fertility Preferences in a Context of Uncertainty. *Population and Development Review*, 44(1):87–116, March 2018. ISSN 00987921. doi: 10.1111/padr.12114.
- Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, September 1974. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.185.4157.1124.
- J. van de Wijgert, N. Padian, S. Shiboski, and C. Turner. Is audio computer-assisted self-interviewing a feasible method of surveying in Zimbabwe? *International Journal of Epidemiology*, 29(5):885–890, October 2000. ISSN 0300-5771.
- Susan Cotts Watkins. Back to Basics: Gender, Social Norms, and the AIDS Epidemic in Sub-Saharan Africa. In David E. Sahn, editor, *The Socioeconomic Dimensions of HIV/AIDS in Africa: Challenges, Opportunities and Misconceptions*, pages 134–162. Cornell University Press, Ithaca, NY, 2011.
- Maria J. Wawer, Ronald H. Gray, Nelson K. Sewankambo, David Serwadda, Xianbin Li, Oliver Laeyendecker, Noah Kiwanuka, Godfrey Kigozi, Mohammed Kiddugavu, Thomas Lutalo, Fred Nalugoda, Fred Wabwire-Mangen, Mary P. Meehan, and Thomas C. Quinn. Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *The Journal of Infectious Diseases*, 191(9):1403–1409, May 2005. ISSN 0022-1899. doi: 10.1086/429411.
- A. A. Weinreb. The limitations of stranger-interviewers in rural Kenya. *American Sociological Review*, 71(6):1014–1039, 2006.
- Parke Wilde and Mark Nord. The Effect of Food Stamps on Food Security: A Panel Data Approach. *Applied Economic Perspectives and Policy*, 27(3):425–432, 2005. ISSN 2040-5804. doi: 10.1111/j.1467-9353.2005.00239.x.
- Daniel John Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98, March 2010. ISSN 1573-6938. doi: 10.1007/s10683-009-9230-z.

Appendix A

Appendix to Chapter 3

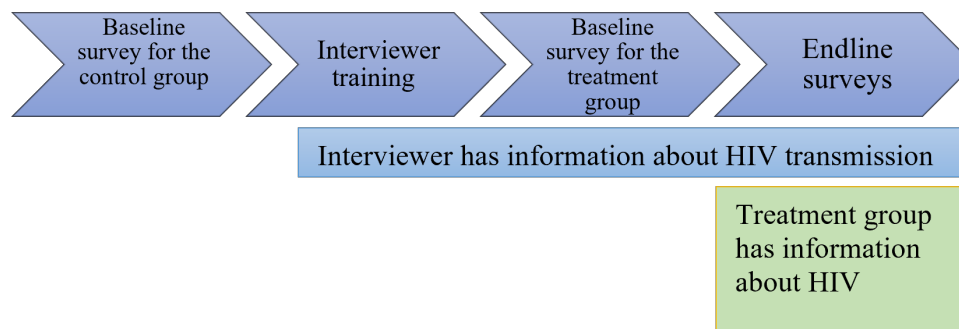


Figure A.1: Timeline of Experiment

Table A.1: HIV Transmission Risk Survey Questions (Female Versions)

- H1a.** If 100 women, who do not have HIV, each have an HIV-positive sex partner for one year, and do not use condoms when having sex, how many of the women do you think will have HIV at the end of the year?
- H2a.** If 100 women, who do not have HIV, each have an HIV-positive sex partner for one year, and do use condoms when having sex, how many of the women do you think will have HIV at the end of the year?
- E1a.** If 100 women, who do not have HIV, each sleep with a man who is HIV positive tonight and do not use a condom, how many of them do you think will have HIV after the night?
- E2a.** If 100 women, who do not have HIV, each sleep with a man who is HIV positive tonight and do use a condom, how many of them do you think will have HIV after the night?

Notes: Survey questions were gender-specific, so men were asked about 100 men and women were asked about 100 women. All survey questions were asked Chichewa (translated versions available upon request).

Table A.2: Balance of Sexual Activity and Demographics

	Control Mean (SD) (1)	Treatment Mean (SD) (2)	Difference (p-value) (3)	N (4)
<u>Panel A: Sexual Activity</u>				
Any Sex in Past Week	0.541 (0.499)	0.507 (0.500)	-0.036 (0.111)	1,292
Sex Acts in Past Week	1.798 (2.471)	1.615 (2.380)	-0.185 (0.155)	1,292
Unprotected Sex Acts in Past Week	1.569 (2.376)	1.471 (2.323)	-0.100 (0.446)	1,292
Sex Partners in Past 30 Days	0.818 (0.498)	0.797 (0.762)	-0.024 (0.515)	1,290
Condoms Acquired in Past 30 Days	4.739 (15.003)	3.530 (11.549)	-1.205 (0.122)	1,288
Years Sexually Active	13.100 (8.279)	13.204 (8.603)	0.117 (0.815)	1,275
Lifetime Sex Partners	3.117 (2.684)	3.557 (4.734)	0.414** (0.042)	1,288
Any Chance They Have HIV	0.344 (0.475)	0.352 (0.478)	0.008 (0.788)	1,277
Overall Sexual Activity Index	0.035 (1.031)	-0.035 (0.968)	-0.072 (0.168)	1,292
Diary Sexual Activity Index	0.028 (0.997)	-0.028 (1.003)	-0.059 (0.266)	1,277
Alternate Sexual Activity Index†	0.018 (1.003)	-0.053 (1.022)	-0.074 (0.149)	1,287
<u>Panel B: Demographics</u>				
Male	0.425 (0.495)	0.436 (0.496)	0.000 (1.000)	1,292
Married	0.829 (0.377)	0.803 (0.398)	-0.025 (0.316)	1,290
Age	29.133 (8.417)	29.589 (8.333)	0.465 (0.339)	1,292
Years of Schooling	5.758 (3.347)	5.858 (3.484)	0.097 (0.723)	1,292
Total Living Children	2.940 (2.128)	3.049 (2.389)	0.113 (0.373)	1,292
Number of People in Household	5.039 (2.237)	4.870 (2.036)	-0.176 (0.254)	1,292
Spending in Past 30 Days (PPP USD)	292.390 (383.593)	293.010 (572.544)	1.698 (0.954)	1,292
Total Assets Owned	4.188 (2.427)	3.937 (2.311)	-0.248 (0.192)	1,292
Christian	0.910 (0.286)	0.927 (0.260)	0.017 (0.472)	1,292
Muslim	0.085 (0.280)	0.060 (0.238)	-0.025 (0.281)	1,292

Notes: This table replicates the balance tests from Kerwin (2019). Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. † Alternate Sexual Activity Index is a PCA index of the first five variables in the table, to check for a potential aggregate balance problem in those variables. p-values are for cluster-adjusted t-tests: * p<0.1; ** p<0.05; *** p<0.01.

Table A.3: Effects of Interviewer Knowledge on Reported Risk Beliefs, Controlling for Interviewer and Sampling Strata Fixed Effects Only

	(1)	(2)	(3)	(4)
	<u>Outcome: HIV transmission risk belief</u>			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment	-0.0736*** (0.0190)	-0.0348*** (0.0128)	-0.0371*** (0.0127)	-0.0867*** (0.0155)
Control-group mean	0.828	0.928	0.123	0.237
Control-group SD	0.264	0.169	0.218	0.281
Observations	1,287	1,282	1,289	1,282

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes

1,292 sexually-active adults who were successfully interviewed at both baseline and endline.

Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table A.4: Effects of Interviewer Knowledge on Reported Risk Beliefs, Controlling for Sampling Strata Fixed Effects Only

	(1)	(2)	(3)	(4)
	<u>Outcome: HIV transmission risk belief</u>			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment	-0.0928*** (0.0188)	-0.0482*** (0.0119)	-0.0271** (0.0114)	-0.0795*** (0.0151)
Control-group mean	0.828	0.928	0.123	0.237
Control-group SD	0.264	0.169	0.218	0.281
Observations	1,289	1,284	1,291	1,284

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes

1,292 sexually-active adults who were successfully interviewed at both baseline and endline.

Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table A.5: Distribution of Religious Denomination and Ethnic Group By Treatment Arm

	Control Mean (SD) (1)	Treatment Mean (SD) (2)	Difference (p-value) (3)	N (4)
<u>Panel A: Religion</u>				
Quadiriya Muslim	0.057 (0.233)	0.048 (0.214)	-0.009 (0.664)	1,291
Sukut Muslim	0.028 (0.165)	0.012 (0.111)	-0.016** (0.045)	1,291
Other Muslim	0.000 (0.000)	0.002 (0.039)	0.002 (0.310)	1,291
Catholic	0.169 (0.375)	0.209 (0.407)	0.039 (0.235)	1,291
Church of Christ	0.194 (0.396)	0.196 (0.397)	0.002 (0.962)	1,291
Anglican	0.019 (0.135)	0.028 (0.165)	0.010 (0.507)	1,291
CCAP	0.126 (0.332)	0.093 (0.290)	-0.034 (0.220)	1,291
Pentecostal	0.144 (0.352)	0.085 (0.279)	-0.058** (0.017)	1,291
Indigenous Christian	0.067 (0.250)	0.028 (0.165)	-0.039*** (0.006)	1,291
Other Christian	0.193 (0.395)	0.289 (0.454)	0.096*** (0.002)	1,291
African Traditional Religion	0.002 (0.039)	0.002 (0.039)	0.000 (0.999)	1,291
Atheist	0.000 (0.000)	0.008 (0.088)	0.008** (0.019)	1,291
Don't Know	0.002 (0.039)	0.002 (0.039)	0.000 (0.992)	1,291
<u>Panel B: Ethnicity</u>				
Nyanja	0.460 (0.499)	0.482 (0.500)	0.020 (0.678)	1,292
Lomwe	0.343 (0.475)	0.394 (0.489)	0.052 (0.337)	1,292
Yao	0.107 (0.309)	0.068 (0.252)	-0.039 (0.156)	1,292
Chewa	0.048 (0.214)	0.031 (0.173)	-0.017* (0.072)	1,292
Other	0.042 (0.200)	0.025 (0.155)	-0.017 (0.111)	1,292

Notes: All regressions control for stratification cell and interviewer fixed effects.

Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table A.6: Literacy Rates in Chichewa by Ethnic Group and Study Arm

	Control Mean (SD)	Treatment Mean (SD)	Difference (p-value)	N
	(1)	(2)	(3)	(4)
Literate and Nyanja	0.329 (0.470)	0.358 (0.480)	0.027 (0.476)	1,289
Literate and Lomwe	0.256 (0.437)	0.290 (0.454)	0.033 (0.406)	1,289
Literate and Yao	0.068 (0.252)	0.050 (0.217)	-0.019 (0.355)	1,289
Literate and Chewa	0.039 (0.193)	0.025 (0.156)	-0.014 (0.134)	1,289
Literate and Other Ethnicity	0.036 (0.186)	0.016 (0.124)	-0.020** (0.026)	1,289

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table A.7: Effects of Interviewer Knowledge on Reported Risk Beliefs, Adding Controls for Religion, Ethnicity, and Literacy

	(1)	(2)	(3)	(4)
	Outcome: HIV transmission risk belief			
	Per Act, Unprotected	Per Year, Unprotected	Per Act, With Condom	Per Year, With Condom
Treatment	-0.081*** (0.019)	-0.041*** (0.013)	-0.038*** (0.014)	-0.084*** (0.016)
Control-group mean	0.828	0.928	0.123	0.235
Control-group SD	0.263	0.169	0.218	0.280
Observations	1,278	1,273	1,280	1,273

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * p<0.1; ** p<0.05; *** p<0.01.

Table A.8: Treatment Effect Spillovers

	(1)	(2)	(3)	(4)
	Outcome: HIV transmission risk belief at endline			
	Main Effect of Treatment	Pooled Spillover Effect of Treatment	Spillover Effects of Treatment by Study Arm	Spillover Effects of Treatment for Control Group Only
Treatment	-0.397*** (0.019)	-0.397*** (0.040)	-0.406*** (0.045)	
Treated Friends		-0.001 (0.013)		-0.017 (0.022)
Control \times (Treated Friends)			-0.012 (0.023)	
Treatment \times (Treated Friends)			0.019 (0.020)	
Total Friends		-0.014 (0.008)		0.012 (0.011)
Control \times (Total Friends)			0.021* (0.011)	
Treatment \times (Total Friends)			-0.004 (0.014)	
Control-group mean	0.829	0.829	0.829	0.829
Control-group SD	0.263	0.263	0.263	0.263
Observations	1,267	1,267	1,267	634

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.9: Collinearity Diagnostics for Age, Years of Schooling, and Years Sexually Active

Panel A: Eigenvalues and Condition Indices

	Eigenvalue	Condition Index
1	3.54	1.00
2	0.37	3.08
3	0.08	6.63
4	0.01	17.85
Det (correlation matrix)	0.21	
Condition Number	17.85	

Panel B: Variance Inflation Factors (VIFs)

Variable	VIF	Sqrt VIF	Tolerance	Rsquared
Age	4.39	2.10	0.23	0.77
Schooling	1.09	1.04	0.92	0.08
Years sexually active	4.53	2.13	0.22	0.78
Mean VIF	3.34			

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.11: Correlations between Schooling Measures and HIV Risk Beliefs in the Control Group

	(1)	(2)
	<u>Outcome: HIV transmission risk belief</u>	
	<u>(per-act, unprotected)</u>	
Years of Schooling	0.009*** (0.003)	0.011*** (0.003)
Controls for other covariates [‡]	N	Y
Control-group mean	0.8278	0.832
Control-group SD	0.264	0.261
Observations	644	629

Notes: All regressions control for stratification cell and interviewer fixed effects. Sample includes 1,292 sexually-active adults who were successfully interviewed at both baseline and endline. Heteroskedasticity-robust standard errors, clustered by village, in parentheses:

* p<0.1; ** p<0.05; *** p<0.01.